BAR Brazilian Administration Review

ANPAD

Thinking Outside the Box

Why Do Management Scholars Avoid Experiments? A Necessary Provocation

João Fernandes Jorge de Siqueira[1], Rafael Barreiros Porto[2], Jonathan Simões Freitas[1]

[1] Universidade Federal de Minas Gerais, Faculdade de Ciências Econômicas, Belo Horizonte, MG, Brazil
[2] Universidade de Brasília, Faculdade de Administração, Contabilidade, Economia e Gestão de Políticas Públicas, Brasília, DF, Brazil

How to cite: Siqueira, J. F. J., Porto, R. B., & Freitas, J. S. (2026). Why do management scholars avoid experiments? A necessary provocation. *BAR-Brazilian Administration Review, 23*(2), e260039.

DOI: https://doi.org/10.1590/1807-7692bar2026260039

Keywords:
empirical research; experimental methods; administrative sciences; organization and management; causal analysis.

JEL Code:
M10

Publication date:
April 15, 2026.

Corresponding author:
João Fernandes Jorge de Siqueira
Universidade Federal de Minas Gerais, Faculdade de Ciências Econômicas
Avenida Presidente Antônio Carlos, n. 6627, Pampulha, CEP 31270-901, Belo Horizonte, MG, Brazil

Funding:
This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brasil) - Finance Code 001.

Conflict of Interests:
The authors stated that there was no conflict of interest.

ABSTRACT

Despite the consolidation of experimental designs as a central standard for causal inference in adjacent fields, experiments remain peripheral in large segments of management research. This article argues that such marginalization is not primarily technical, but epistemic and institutional. It reconstructs six recurrent objections — complexity, external validity, feasibility, theory reduction, non-manipulability, and ethical scope — that structure skepticism toward experimentation and shows how they normalize the substitution of statistical sophistication for design-based identification. The analysis suggests that resistance to experiments reflects entrenched evaluative norms about what counts as rigor and relevance, rather than demonstrated methodological inadequacy. To move beyond dichotomous debates, the article introduces a simple evaluative framework structured along two dimensions: causal ambition and organizational embeddedness. By conceptualizing experimentation as a continuum, the framework aligns the strength of causal claims with the inferential capacities of different designs, making trade-offs explicit rather than implicit. The central contribution is disciplinary rather than technical: repositioning experimentation as a reference point for transparent causal reasoning. The article concludes by calling for greater alignment between causal claims and research design, emphasizing inferential discipline as a condition for credible knowledge in management research.

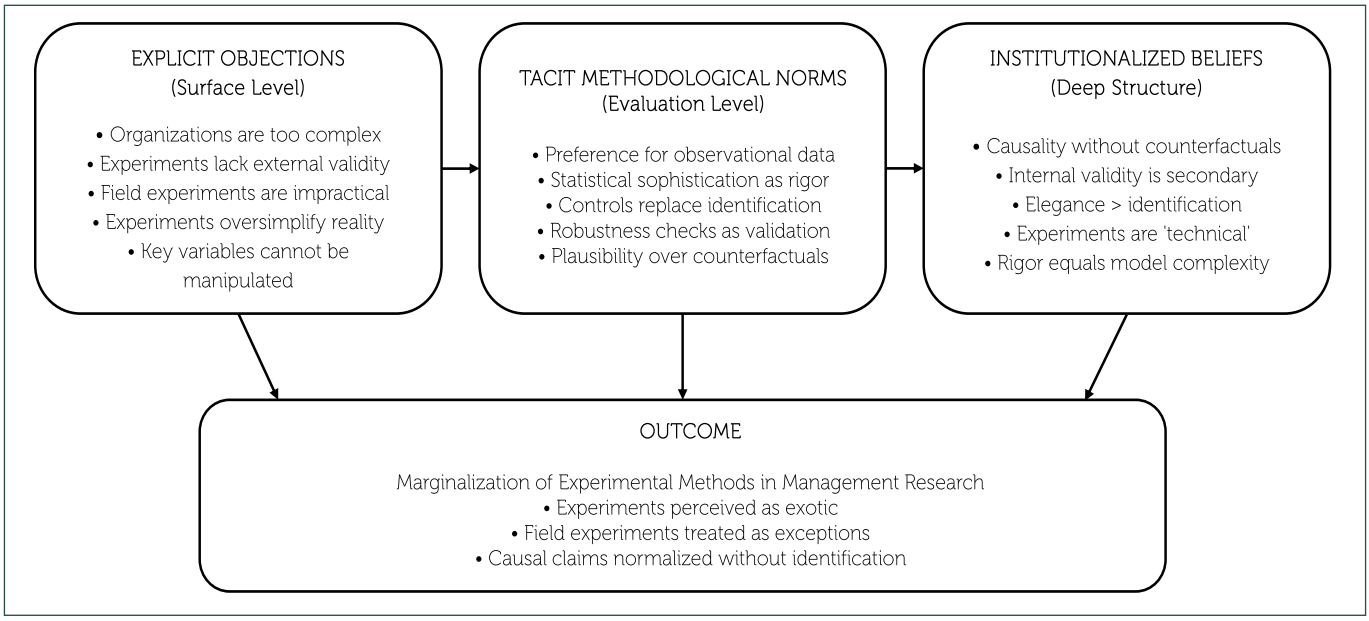
INTRODUCTION

Despite decades of methodological advances and the consolidation of experimental designs as a central standard for causal inference in adjacent fields, experiments remain peripheral in large segments of management research. This marginality is not accidental. It is sustained by a persistent grammar of objections that normalizes the substitution of statistical sophistication and extensive controls for design-based identification. As a result, causal claims are frequently advanced without credible counterfactual reasoning or exogenous variation. In contrast, the experimental tradition grounds inference in design rather than post hoc adjustment, treating identification, not plausibility or elegance, as the core criterion of causal credibility.

This article argues that resistance to experimentation reflects more than technical constraints. Methodological preferences operate as signals of disciplinary belonging, shaping what is recognized as rigorous, relevant, and publishable knowledge. To examine this pattern, the article reconstructs six recurrent objections that structure the field's skepticism toward experiments and analyzes the epistemic costs of their normalization. By organizing these objections, the article clarifies how they delimit causal ambition while framing their limits as methodological maturity.

The primary objective is epistemological and disciplinary: to challenge naturalized assumptions about causality and rigor in management research. Two secondary contributions follow. First, the article systematizes the field's recurrent objections to experimentation and makes their trade-offs explicit. Second, it introduces a simple evaluative framework that links degrees of experimentalization to the causal claims research designs can credibly sustain. Together, these contributions reposition experimentation not as a methodological niche, but as a reference point for more transparent causal reasoning.

Figure 1 illustrates how explicit critiques of experimental methods (at the surface level) are translated into tacit methodological norms used in peer evaluation and, over time, sedimented as institutionalized beliefs about rigor and sophistication in management research.



Source: Developed by the authors.

Figure 1. The temporal sedimentation of objections to experimental methods in management research.

This figure illustrates how explicit critiques of experimental methods, initially articulated as surface-level methodological objections, are gradually translated into tacit evaluation norms within peer review and doctoral training. Over time, these norms become sedimented as institutionalized beliefs about rigor, relevance, and sophistication in management research. What begins as a contestable methodological concern evolves into a taken-for-granted epistemic standard, shaping publication incentives and legitimizing causal claims grounded in plausibility, statistical complexity, and post hoc adjustment rather than in design-based identification.

Objection 1: "Organizations are too complex to be experimented on"

Perhaps the most common criticism is that organizations are complex, open systems, characterized by

interdependence, emergence, and contextual specificity. From this perspective, experiments are accused of oversimplifying reality, artificially isolating variables, and ignoring the richness of organizational life.

This argument is intuitively appealing, but epistemically misleading.

In practice, appeals to complexity often function as a license for post hoc causal storytelling. When everything is treated as contingent, context-dependent, and interconnected, almost any observed outcome can be retrospectively explained. Competing explanations are not eliminated but accumulated. Complexity thus becomes a rhetorical shield: causal claims are advanced not because alternative explanations have been ruled out, but because they can always be accommodated after the fact. If everything matters, then nothing is decisively tested. This pattern echoes classic critiques of unfalsifiable theorizing, in which explanations survive precisely because no research design forces them to fail (Bacharach, 1989; Merton, 1968).

This logic reveals a deeper disciplinary tendency toward the valorization of descriptive richness at the expense of explanatory discipline. Thick descriptions, nuanced narratives, and context-sensitive accounts undoubtedly enrich our understanding of organizational life, but description is not explanation. Without controlled variation, rich descriptions coexist with causal claims; they do not justify them. As long emphasized in causal methodology, explanation requires designs that can rule out plausible alternatives, not narratives that can indefinitely absorb them (Holland, 1986).

Complexity, therefore, does not invalidate experimentation; it raises the bar for design. Field experiments, by definition, operate within complexity rather than against it. They do not seek to recreate organizations in sterile environments, but to introduce controlled, exogenous variation into naturally occurring organizational processes.

Under conditions of high complexity, purely observational designs face their greatest challenges, as confounding, simultaneity, and selection effects become harder, not easier, to address. The greater the complexity of the organizational setting, the higher the epistemic premium on research designs capable of generating exogenous variation. In such contexts, causal inference cannot rely solely on plausibility, narrative coherence, or statistical adjustment; it requires designs that make causal contrasts observable rather than assumable (Rosenbaum, 2002).

Objection 2: "Experiments lack external validity"

A second, closely related critique is that experiments produce results that do not generalize to real organizations. Although this concern historically emerged in response to laboratory experiments, it has migrated, largely by methodological inertia, to virtually all experimental designs, including field experiments. As a result, experiments are often treated as inherently artificial, regardless of whether they are conducted in classrooms, online platforms, or inside functioning organizations (Harrison & List, 2004).

This critique is increasingly outdated.

Field experiments were developed precisely to address concerns about realism and generalizability by embedding randomization into real-world organizational contexts. When interventions are implemented with actual managers, employees, customers, technologies, and institutional constraints, ecological validity is not a limitation; it is a defining feature of the design. In many cases, such studies exhibit higher realism and contextual embedding than survey-based or archival research that relies on self-reports, proxies, or administrative records detached from decision-making processes (Gottschalk & Moffitt, 2009).

A more productive way to frame this issue is to distinguish among realism, generalizability, obtrusiveness, and precision. Laboratory experiments typically maximize precision and control, often at the expense of realism. Field experiments trade some precision for substantially higher realism and, frequently, stronger claims about external validity. Observational designs, while minimally obtrusive and often broad in scope, rely heavily on assumptions for identification. As a result, discussions of external validity are epistemically hollow unless preceded by credible identification of causal effects (Cook, 2015).

The absence of experimental control in surveys or archival studies does not guarantee external validity, because generalizing correlations without credible identification amplifies uncertainty rather than knowledge (Shadish et al., 2002).

External validity is not the opposite of internal validity, nor a substitute for it. In management research, claims about generalizability are only as credible as the causal contrasts on which they are built. Field experiments do not eliminate the problem of external validity, but they anchor it in empirically identifiable effects rather than inferential hope.

Objection 3: "Experiments are impractical and too costly"

Management scholars frequently argue that experiments, particularly field experiments, are logistically infeasible, prohibitively expensive, time-consuming, and dependent on organizational access that is difficult to secure.

This objection is partially true and strategically exaggerated.

Field experiments do require negotiation, trust-building, and sustained collaboration with organizations. They demand careful planning, ethical sensitivity, and alignment between research questions and organizational routines. However, these requirements are not unique to experiments. High-quality qualitative studies, ethnographies, and longitudinal case analyses impose comparable, and often greater, demands in terms of time, access, and relational work, yet they are rarely dismissed as impractical. The difference is not feasibility, but methodological familiarity (Van de Ven, 2007).

More fundamentally, this objection reflects a gap in researcher training. Experimentation often appears costly and unmanageable not because it is inherently so, but because management education has underinvested in training researchers in intervention design, randomization, and implementation under organizational constraints (Aguinis & Bradley, 2014).

Moreover, many organizations already embed A/B testing, randomized rollouts, and pilot programs into routine decision processes, suggesting that experimentation is more viable in practice than often assumed (Kohavi et al., 2020).

Ironically, while organizations rely on experimental logic to guide managerial action, academic management research has lagged behind, constrained less by access or cost than by a persistent lack of methodological imagination (Gottschalk & Moffitt, 2009). The practicality of experiments thus depends not only on organizational conditions, but on the field's willingness to recognize experimentation as an endogenous feature of contemporary organizations.

Objection 4: "Experiments reduce theory to narrow causal claims"

Another, more subtle resistance concerns theory. Experiments are often accused of producing 'small,' 'local,' or 'incremental' findings, causal effects that are context-bound and theoretically unambitious.

This critique confuses precision with triviality.

In many cases, the problem is not that experimental findings are narrow, but that prevailing theories in management are written in ways that render them effectively irrefutable. When theories are elastic enough to accommodate almost any outcome, empirical precision becomes a threat rather than a virtue. Long-standing critiques of organizational theory warn that ideas often survive not because they are powerful, but because they are insulated from decisive empirical tests (Bacharach, 1989; Merton, 1968).

Experiments do not replace theory; they discipline it. By requiring explicit articulation of mechanisms, boundary conditions, and counterfactual comparisons, experimental designs impose a level of theoretical clarity that observational approaches can often evade. Vague constructs, underspecified causal chains, and post hoc rationalizations are far more difficult to sustain when causal contrasts are built into the design. In this sense, experiments do not weaken theory; they make theoretical failure visible, and thereby enable theoretical progress (Bacharach, 1989).

Not every experiment aspires to universal causality. Identifying what works, for whom, and under which conditions is not a theoretical weakness; it is a marker of theoretical maturity.

Objection 5: "Causality in organizations is not manipulable"

A further belief is that key organizational phenomena, such as culture, leadership, strategy, and institutions, cannot be manipulated ethically or practically, rendering experimentation inappropriate or superficial.

This objection rests on a false dichotomy.

Field experiments rarely attempt to manipulate abstract constructs directly. Instead, they intervene on the concrete levers through which organizational phenomena are enacted: decision rules, incentive structures, defaults, feedback mechanisms, information architectures, and process designs. These are not peripheral details; they are the operational substrates through which strategy, leadership, and culture materialize in organizational practice (Gerring, 2007).

When the phenomenon of interest is abstract, the experimental design operates not on labels, but on components. Refusing to experiment in such contexts reflects not a principled limitation, but a reification of constructs, treating them as immutable entities rather than as patterns of action produced by organizational design. By decomposing 'large' phenomena into manipulable components, experimental methods render them analytically tractable without trivializing their significance.

Objection 6: "Experiments cannot address strategic or ethical phenomena"

A final resistance concerns strategy and ethics, with experiments portrayed as suitable for micro-level behaviors but ill-equipped to address strategic choice, power, or ethical judgment. This view rests on a narrow conception of experimentation. Experimental designs do not manipulate 'strategy' or 'ethics' in the abstract; they operate on decision architectures, information structures, incentives, defaults, and governance rules,

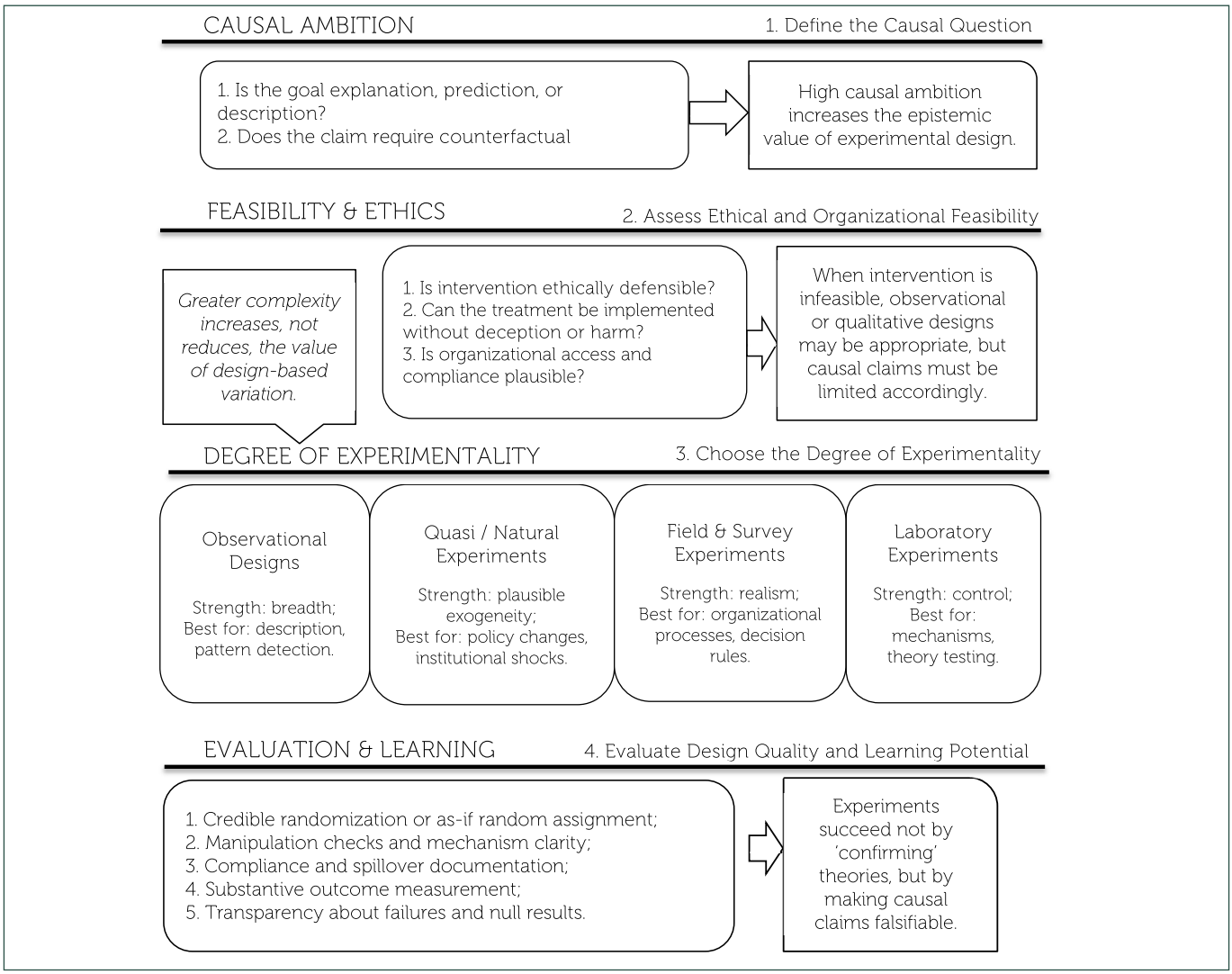
through which strategic and ethical outcomes are enacted (Cartwright, 2007; Gerring, 2007). Many consequential organizational decisions are already shaped by implicit experimentation in algorithmic systems and digital platforms (Kohavi et al., 2020). By requiring ex ante specification of mechanisms, outcomes, and counterfactuals, experiments can make strategic and ethical trade-offs more explicit rather than less, reinforcing causal clarity in domains where consequences are significant (Hausman, 2012).

It is intended not to constrain methodological diversity, but to render it transparent, transforming experimentation from a methodological niche into a visible standard against which causal reasoning in management research can be more openly evaluated.

A SYNTHESIS, A FRAMEWORK, AND A CALL TO ACTION

Taken together, the objections examined throughout this article reveal less about the intrinsic limitations of experimental methods and more about the epistemic culture of management research. Across domains (complexity, external validity, feasibility, theory, manipulability, and ethics), experiments are portrayed as problematic not because of demonstrated failure, but because they unsettle established hierarchies of evidence. The claim advanced here is epistemological: when strong causal explanations are at stake, designs that deliberately introduce and control variation, particularly well-executed field experiments, provide the most credible basis for causal identification, provided their assumptions and constraints are made explicit. The issue is not methodological pluralism, but the absence of a shared language for calibrating causal claims to design properties.

To move the debate beyond dichotomies, this article proposes a simple framework (Figure 2) of degrees of experimentality in management research. Rather than treating experimentation as a binary methodological choice, the framework conceptualizes it as a continuum defined by the extent to which causal variation is deliberately introduced and controlled by design.



Source: Developed by the authors.

Figure 2. A decision framework for experimental research in management.

To address this gap, the framework proposed here is structured along two analytically independent dimensions: causal ambition, defined as the strength and scope of the causal claims a study seeks to advance; and organizational embeddedness, defined as the extent to which an intervention is implemented within naturally occurring organizational processes and decision structures. The framework's core purpose is calibration: high causal ambition unsupported by design-based sources of variation results in overclaiming, whereas high embeddedness without credible identification produces contextually rich but causally indeterminate findings.

Within this two-dimensional space, laboratory and vignette experiments typically combine strong control over treatment assignment with lower embeddedness, facilitating mechanism testing under simplified conditions; natural experiments leverage exogenous shocks to approximate identification with limited researcher control; and field experiments integrate deliberate variation into ongoing organizational settings, increasing contextual relevance while introducing implementation constraints. Rather than ranking these designs, the framework clarifies the inferential trade-offs each entails and aligns the credibility of causal claims with the structure of the research design.

Applying the framework: An illustrative diagnostic example

To illustrate the diagnostic logic of the framework, consider a stylized and representative study in the leadership and creativity literature. The article examines whether participatory leadership increases employee creativity using multi-organization survey data, regression models with extensive controls, robustness checks, and mediation analyses, and concludes that participatory leadership causally enhances creative performance.

Viewed through the framework, the study advances strong causal language ('leads to,' 'drives,' 'produces') without a design-based source of variation capable of supporting the implied counterfactual comparison. Its identification strategy rests primarily on statistical adjustment and theoretical plausibility. At the same time, the study benefits from a rich organizational context (real firms, leaders, and employees), which strengthens its substantive relevance but does not eliminate observational equivalence among competing explanations such as self-selection or reverse causality.

The framework does not invalidate the contribution; it recalibrates it. Under this lens, the study offers robust associational evidence consistent with participatory leadership theory, refines constructs, and motivates subsequent causal testing. What it cannot credibly sustain is a strong causal claim. Approached instead through randomized leadership training across comparable units, the same question could support higher causal ambition, albeit with trade-offs in scope and implementation. A natural experiment exploiting exogenous institutional change would occupy an intermediate position. The example demonstrates the framework's function: to discipline the alignment between causal claims and design properties by making inferential trade-offs explicit rather than implicit.

Design-specific advantages: What each experimental approach does best

Building on this diagnostic logic, the framework allows experimental and quasi-experimental designs to be evaluated not by their limitations, but by their distinct epistemic advantages. Each design excels under specific inferential and organizational conditions. In concise terms, laboratory experiments maximize control and mechanism isolation; vignette and survey-based experiments expand theoretical reach and scalability across populations; natural experiments strengthen identification by leveraging exogenous institutional or policy variation; and field experiments combine credible causal inference with real organizational behaviors and outcomes. These designs are therefore complements rather than substitutes, aligning differently with configurations of causal ambition and organizational embeddedness. Mature research programs often move across them sequentially, refining mechanisms under control, probing scope, exploiting institutional variation, and ultimately assessing real-world impact.

Taken together, this application reinforces the article's central claim: methodological rigor in management research should be evaluated not by stylistic sophistication or dominant conventions, but by the coherence between causal claims, design properties, and organizational realities. The framework does not ask whether a study is experimental; it asks whether its inferential ambition is justified, reframing experimentation as a shared evaluative language for transparent causal reasoning.

LAST THOUGHTS

Normalizing experimental reasoning in administration requires a shift in research practice: design must precede estimation. Causal claims should be anchored in explicit identification strategies rather than reinforced ex post by statistical sophistication. This implies treating intervention design, credible randomization or as-if random assignment, attention to compliance and spillovers, and clear articulation of counterfactuals as core

components of theorizing. The objective is not methodological conversion, but tighter alignment between causal ambition and design capacity.

Implications extend to graduate training and editorial governance. Doctoral programs should teach experimental, quasi-experimental, qualitative, and observational designs comparatively, cultivating design literacy and explicit reasoning about identification. Review processes, in turn, should foreground transparent criteria for evaluating causal claims, clarifying expectations about identification, treatment integrity, and inferential limits, so that methodological assessment becomes design-centered rather than convention-driven. A shared evaluative language reduces arbitrariness and strengthens cumulative learning.

Experiments will not replace other methods, but their marginalization carries costs. When causal ambition exceeds design support, the field accumulates persuasive yet fragile claims. The core provocation of this article is simple: credible explanation requires inferential discipline. Embracing experimental reasoning is less about methodological preference than about accepting empirical risk and allowing causal claims to be constrained by design rather than sustained by plausibility alone.

REFERENCES

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods, 17*(4), 351-371. https://doi.org/10.1177/1094428114547952

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review, 14*(4), 496–515. https://doi.org/10.5465/amr.1989.4308374

Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge University Press.

Cook, T. D. (2015). Generalizing causal knowledge in the policy sciences: External validity as a research objective. *Policy Studies Journal, 43*(4), 527–549. https://doi.org/10.1111/psj.12123

Gerring, J. (2007). *Case study research: Principles and practices*. Cambridge University Press.

Gottschalk, P., & Moffitt, R. (2009). The rising instability of U.S. earnings. *Journal of Economic Perspectives, 23*(4), 3–16. https://doi.org/10.1257/jep.23.4.3

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature, 42*(4), 1009–1055. https://doi.org/10.1257/0022051043004577

Hausman, D. M. (2012). *Preference, value, choice, and welfare*. Cambridge University Press.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945-960. https://doi.org/10.1080/01621459.1986.10478354

Kohavi, R., Tang, D., Xu, Y., & Walker, T. (2020). *Trustworthy online controlled experiments: A practical guide to A/B testing*. Cambridge University Press.

Merton, R. K. (1968). *Social theory and social structure*. Free Press.

Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). Springer.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Van de Ven, A. H. (2007). *Engaged scholarship: A guide for organizational and social research*. Oxford University Press.

Authors

João Fernandes Jorge de Siqueira

Universidade Federal de Minas Gerais, Faculdade de Ciências Econômicas

Avenida Presidente Antônio Carlos, n. 6627, Pampulha, CEP 31270-901, Belo Horizonte, MG, Brazil

jfjsiqueira@yahoo.com

Rafael Barreiros Porto

Universidade de Brasília, Faculdade de Administração, Contabilidade, Economia e Gestão de Políticas Públicas

Campus Universitário Darcy Ribeiro, Asa Norte, CEP 70910-900, Brasília, DF, Brazil

rafaelporto@unb.br

Jonathan Simões Freitas

Universidade Federal de Minas Gerais, Faculdade de Ciências Econômicas

Avenida Presidente Antônio Carlos, n. 6627, Pampulha, CEP 31270-901, Belo Horizonte, MG, Brazil

jonathansf@face.ufmg.br

Authors' contributions

1st author: conceptualization (equal), formal analysis (equal), investigation (equal), methodology (equal), validation (equal), visualization (equal), writing - original draft (equal), writing - review & editing (equal).

2nd author: conceptualization (equal), supervision (equal), validation (equal), writing - original draft (equal), writing - review & editing (equal).

3rd author: conceptualization (equal), formal analysis (equal), methodology (equal), supervision (equal), validation (equal), writing - original draft (equal), writing - review & editing (equal).