

Pesquisa Empírica Quantitativa em Contabilidade Gerencial: uma Proposta de Tipologia e Implicações para Validade Interna versus Validade Externa

Andson Braga de Aguiar

<http://orcid.org/0000-0003-4034-4134>

Daniel Magalhães Mucci

<https://orcid.org/0000-0002-0658-1470>

Myrna Modolon Lima

<https://orcid.org/0000-0003-2084-481X>

Resumo

Objetivo: São dois os objetivos deste estudo. Primeiramente, propomos uma tipologia de pesquisa empírica quantitativa em contabilidade gerencial baseada em duas características do desenho de pesquisa: grupo de controle e representatividade da amostra. Em segundo lugar, discutimos as implicações de cada método considerando as vantagens e desvantagens (*trade offs*) entre a validade interna e validade externa.

Método: Com base em estudos metodológicos anteriores, desenvolvemos uma tipologia com oito métodos empíricos quantitativos.

Resultados: Propomos oito métodos empíricos quantitativos para estudos em contabilidade gerencial baseados em duas características do desenho de pesquisa: (1) experimento de laboratório, (2) experimento *crowdsourcing*, (3) experimento de campo, (4) experimento natural, (5) pesquisa de entidade única, (6) pesquisa documental exclusiva (*proprietary archival study*), (7) levantamento em larga escala e (8) pesquisa documental pré-estruturada (*pre-structured archival study*). Além disso, comparamos de forma crítica as vantagens e desvantagens e discutimos as implicações desses métodos para a validade interna e validade externa.

Contribuições: Esse estudo apresenta duas contribuições. Primeiramente, a tipologia proposta pode ajudar o pesquisador júnior da área de contabilidade gerencial a se familiarizar com os métodos empíricos disponíveis, alguns dos quais ainda incipientes no Brasil. Em segundo lugar, se estudo mostra que a escolha de um método empírico normalmente implica benefícios para um tipo de validade (por ex.: validade interna) em detrimento de outro tipo de validade (por ex.: validade externa). Alegações de causalidade e a generalização dos resultados dependem de qual tipo de validade é priorizado e dos métodos usados para aumentar a validade geral dos resultados de um estudo.

Palavras chave: Métodos empíricos quantitativos; Contabilidade gerencial; Grupo de controle; Representatividade da amostra; Tipos de validade.

Editado em Português e Inglês. Versão original em Português.

Recebido em 16/8/2022. Aceito em 19/8/2022 por Dr. Gerlando Augusto Sampaio Franco de Lima (Editor). Publicado em 14/10/2022.
Organização responsável pelo periódico: Abracicon.

1. Introdução

Não é incomum que pesquisadores na área de contabilidade gerencial (CG) nos estágios iniciais de suas carreiras, e eventualmente pesquisadores mais experientes, façam afirmações que não podem ser apoiadas pelo método empírico usado para projetar o estudo. Por exemplo, pesquisadores podem ser tentados a fazer alegações causais em estudos que não atendem às condições necessárias para tal quando não há explicações alternativas plausíveis para o efeito além da causa (Shadish, Cook, & Campbell, 2002). Da mesma forma, um pesquisador pode ser tentado a fazer afirmações sobre a generalização dos resultados, mesmo que a seleção aleatória, uma estratégia importante para fornecer evidências para generalização, raramente seja o caso (Trochim, Donnelly e Arora, 2016).

Argumentamos que tais afirmações imprecisas são feitas porque os pesquisadores em CG no Brasil, particularmente pesquisadores iniciantes, não estão familiarizados com os métodos empíricos disponíveis e, conseqüentemente, com as implicações dos diferentes métodos em termos de validade. De fato, estudos anteriores mostram consistentemente que a gama de métodos empíricos utilizados por pesquisadores em CG no Brasil é bastante restrita, com a prevalência de levantamentos em larga escala (por ex.: Aguiar, 2018; Frezatti, Aguiar, Wanderley, & Malagueno, 2015). Enquanto os pesquisadores brasileiros na área de CG não se familiarizarem com os métodos disponíveis, eles podem não estar cientes das vantagens e desvantagens relativas de cada método em termos de validade da pesquisa.

Além disso, embora o pesquisador júnior possa acessar os métodos empíricos disponíveis em livros de metodologia (Smith, 2022; Saunders, Lewis, & Thornhill, 2019), esses livros fornecem apenas uma visão limitada por pelo menos duas razões. A primeira é que os livros de metodologia de pesquisa discutem amplamente cada método empírico disponível, como experimento, documental e levantamento, sem considerar que cada método pode ser desmembrado em projetos de pesquisa mais específicos, apresentando tanto benefícios como ameaças à validade. Por exemplo, estudos documentais podem usar dados pré-estruturados ou proprietários (Moers, 2006). Em segundo lugar, esses livros geralmente se concentram nos protocolos a serem seguidos por cada método de pesquisa mais abrangente, carecendo de uma discussão mais específica e crítica sobre as implicações para a estrutura de validade.

A falta de conhecimento sobre os métodos empíricos disponíveis e suas implicações para a validade da pesquisa pode ter conseqüências importantes para o avanço de um campo de pesquisa. Por um lado, a falta de conhecimento sobre as diversas opções de métodos disponíveis pode restringir o escopo das questões de pesquisa que os pesquisadores de CG investigam, bem como ignorar como diferentes métodos empíricos podem ser complementares na abordagem de um mesmo problema de pesquisa (Bloomfield, Nelson, & Soltes, 2016). Por outro lado, a falta de conhecimento sobre as implicações dos métodos disponíveis para a estrutura de validade pode levar o pesquisador de CG a fazer afirmações não fundamentadas, particularmente, em termos de relações causais e da generalização dos resultados.

Dada a importância de aumentar a abrangência dos métodos empíricos utilizados por pesquisadores juniores no Brasil, esse estudo tem um objetivo duplo. Primeiramente, propomos uma tipologia de pesquisa empírica quantitativa em CG, visto que a pesquisa quantitativa é predominante na pesquisa contábil (Hesford, Lee, Van der Stede, & Young, 2006; Nascimento, Junqueira, & Martins, 2010; Aguiar, 2018). Além disso, os critérios de validade para estudos quantitativos diferem em tipo e importância daqueles usados em estudos qualitativos. O foco na CG se dá porque uma tipologia já havia sido proposta para apoiar os pesquisadores na área de contabilidade a selecionar métodos apropriados com base nas tarefas de coleta de dados (Bloomfield et al., 2016). Embora a tipologia de Bloomfield et al. (2016) seja útil, esta não considera as singularidades que os pesquisadores em CG enfrentam no processo de seleção de um método empírico específico a ser utilizado no projeto de pesquisa. Por exemplo, ao selecionar uma pesquisa, os pesquisadores na área de CG precisam determinar em detalhes o nível de análise e a representatividade da amostra (Van der Stede, Young, & Chen, 2006). A tipologia proposta aqui se baseia em duas características do desenho de pesquisa: grupo de controle e representatividade da amostra. Com base nesses critérios, discutimos oito métodos diferentes: (1) experimento de laboratório, (2) experimento de *crowdsourcing*, (3) experimento de campo, (4) experimento natural, (5) pesquisa de entidade única (6) pesquisa documental exclusiva (*proprietary archival study*), (7) levantamento em larga escala (*survey*) e (8) pesquisa documental pré-estruturada.

Em segundo lugar, discutimos as principais implicações da tipologia proposta para o desenho de pesquisa em termos de validade, a qual se refere à “verdade aproximada de uma inferência” (Shadish et al., 2002, p. 34) e a validade dos resultados da pesquisa é inevitavelmente afetada pelo desenho de pesquisa (Dyckman & Zeff, 2014). Cada método empírico afeta de maneira diferente os tipos de validade. Portanto, as vantagens e desvantagens (*trade-offs*) entre os tipos de validade devem ser consideradas. Embora existam quatro tipos de validade, nos concentramos na validade interna e validade externa por dois motivos. Primeiramente, a validade interna é fundamental para alegações causais, enquanto a validade externa é central para afirmações sobre a generalização de resultados (Shadish et al., 2002). Além disso, cada método empírico impõe vantagens e desvantagens para a validade interna e validade externa (Luft & Shields, 2014; Roe & Just, 2009).

A contribuição desse estudo se apresenta de duas formas. Primeiramente, propomos uma tipologia de pesquisa quantitativa em CG com base em duas características fundamentais do desenho de pesquisa que pode expandir o conjunto de ferramentas disponível para que o pesquisador determine o desenho de seu estudo. Por exemplo, destacamos que a escolha de levantamento (*survey*) exige que os pesquisadores na área de CG façam uma escolha posterior relacionada à unidade de observação, ou seja, decidir entre levantamento em grande escala ou de entidade única. Os métodos empíricos incluídos na tipologia proposta podem ajudar o pesquisador júnior na área de CG a se familiarizar com as alternativas disponíveis, algumas das quais ainda incipientes no Brasil, como os desenhos experimentais (Aguiar, 2017; Nascimento et al., 2010). Em segundo lugar, a discussão sobre as implicações da tipologia proposta fornece *insights* sobre a escolha entre os tipos de validade, assim como as vantagens e desvantagens de cada um, e que os pesquisadores na área de CG devem prestar atenção ao selecionar um determinado método empírico. Por exemplo, um método empírico que inclui grupo de controle pode favorecer a validade interna, mas ao mesmo tempo impor desafios em termos de validade externa. Em outras palavras, a escolha de um determinado método empírico implica uma escolha simultânea dos tipos de validade que são favorecidos no estudo ao mesmo tempo em que impõe desafios ao pesquisador.

2. Validade em Pesquisa na Área de Contabilidade Gerencial e Tipologia Proposta

2.1 Validade Interna e Alegações Causais: Grupo de Controle

Existem três tipos principais de estudos de pesquisa cumulativa (Trochim et al., 2016): i) estudos descritivos, que se concentram em documentar quais são as principais características de uma população ou fenômeno; ii) estudos relacionais, que focam na relação entre duas ou mais variáveis e iii) estudos causais, que determinam se uma ou mais variáveis independentes causam uma ou mais variáveis dependentes. A disposição para estabelecer relações causais e fazer tais alegações é um dos principais objetivos dos estudos quantitativos em CG (Van der Stede, 2014). A principal preocupação em estudos causais é aumentar a validade interna, ou seja, “A validade das inferências sobre se a covariação observada entre A (o tratamento presumido) e B (o resultado presumido) reflete uma relação causal de A para B, considerando como essas variáveis foram manipuladas ou medidas” (Shadish et al., 2002, p. 38).

Três condições devem ser atendidas para que um projeto de pesquisa permita alegações causais: a causa precede o efeito, a causa está relacionada ao efeito e não há uma explicação alternativa plausível para o efeito, além da causa (Shadish et al. al., 2002; Luft & Shields, 2014). O pesquisador pode usar argumentos teóricos e/ou variáveis defasadas para afirmar que a causa precede o efeito e também usar diferentes ferramentas estatísticas para mostrar que a causa está relacionada ao efeito. O principal desafio com alegações causais é garantir a ausência de explicações alternativas plausíveis para a suposta relação causal.

A forma mais eficaz de eliminar explicações alternativas é saber qual teria sido o efeito se a causa não estivesse presente, ou seja, criando um contrafactual (Floyd & List, 2016). A criação de um contrafactual é uma escolha de desenho que envolve a presença de um grupo de controle (ou linha de base). Em outras palavras, um **grupo de controle** cria uma inferência contrafactual útil, essencial para pesquisas em que o interesse é estabelecer relações causais (Lonati, Quiroga, Zehnder, & Antonakis, 2018).

A presença de um grupo de controle é a marca dos estudos experimentais (Shadish, Cook, & Campbell, 2002; Trochim, Donnelly, & Arora, 2016), de modo que essa escolha permite separar os estudos experimentais dos não experimentais. Como contrafactual, um grupo de controle representa um grupo de entrevistados/participantes que são comparáveis ao grupo de tratamento (experimental) em todos os aspectos possíveis, sendo a principal diferença o grupo de controle não estar exposto ao tratamento/manipulação (Oehlert, 2003). Por exemplo, em um estudo examinando o efeito de recompensas na motivação dos funcionários, o grupo de tratamento (controle) incluiria participantes que (não) recebem tais recompensas. Portanto, nesse desenho de pesquisa, o pesquisador compararia a motivação dos funcionários alocados no grupo de tratamento e no grupo de controle, usado como contrafactual, e examinaria se o fato de receber a recompensa afeta ou não a motivação desses funcionários.

Em geral, o principal benefício de incluir um grupo de controle é que os pesquisadores podem mitigar a probabilidade de explicações alternativas para uma relação causal e, assim, melhorar a validade interna e fazer alegações causais com mais confiança (Floyd & List, 2016).

2.2 Validade Externa e Alegações de Generalização: Representatividade da Amostra

Os pesquisadores podem coletar dados primários e/ou secundários para abordar a questão de pesquisa. Mas de qualquer forma, devem estabelecer o processo de seleção de unidades de uma população de interesse, sejam indivíduos, grupos, subunidades ou organizações. Portanto, os pesquisadores devem estabelecer o processo de amostragem que lhes permita generalizar os resultados da amostra para a população da qual as unidades foram selecionadas (Speklé & Widener, 2018). Generalizações são desejáveis e comuns em estudos contábeis (Dyckman & Zeff, 2014), mas para isso, os pesquisadores devem aumentar a validade externa, ou seja, “a extensão em que uma relação causal se mantém sobre as variações nas pessoas, ambientes, tratamentos e resultados” (Shadish et al., 2002, p. 83).

Existem duas abordagens principais para selecionar uma amostra: probabilística e não probabilística. A principal diferença entre as duas abordagens é que a amostragem probabilística envolve amostragem aleatória, em que cada unidade da população tem a mesma chance de ser selecionada (Dyckman & Zeff, 2014). A melhor abordagem para fazer afirmações de generalização dos resultados é a probabilística que, no entanto, raramente é viável (Speklé & Widener, 2018). Alternativamente, pode-se usar uma abordagem chamada modelo de similaridade proximal que permite generalizações da amostra observada para outras amostras com base no grau em que as outras amostras são semelhantes à amostra observada (Trochim et al., 2016).

Independentemente da abordagem de amostragem, o desafio é obter **amostras representativas** para que seja possível fazer generalizações. Amostras representativas permitem a apresentação de evidência de validade externa e, dessa forma, a generalização dos resultados. Dada a dificuldade em obter amostras aleatórias, os pesquisadores de CG normalmente obtêm amostras representativas através de amostras heterogêneas, em que as unidades são escolhidas propositalmente para refletir diversidade em importantes dimensões pré-definidas (Shadish et al., 2002). Por exemplo, em um estudo examinando o efeito dos tipos de recompensa no comportamento míope de executivos, pode-se usar dados pré-estruturados de várias organizações, ou seja, dados coletados e armazenados antes do início da pesquisa (Das, Jain, & Mishra, 2016; Moers, 2006), e que estão disponíveis publicamente através de plataformas como a Economática e o Compustat. Outro exemplo seria um estudo examinando os efeitos de sistemas de controle em empresas familiares sobre o comportamento de não membros da família. Amostras representativas podem ser recrutadas através de questionários enviados a um grande número de empresas familiares. Amostras representativas também podem ser obtidas através do uso de base de dados proprietários, ou seja, a base de dados é confidencial e pode ser acessada somente se o proprietário dos dados autorizar o acesso (Moers, 2006). Exemplos de dados proprietários são pesquisas de terceiros (por exemplo, empresas de consultoria) e dados internos de empresas (Das et al., 2016; Moers, 2006).

Em geral, o principal benefício de amostras representativas é aprimorar a validade externa para melhor fundamentar afirmações relacionadas à generalização de resultados (Trochim et al., 2016).

2.3 Tipologia Proposta

Com base nesses dois critérios – grupo de controle e representatividade da amostra –, propomos uma tipologia com oito alternativas de métodos empíricos quantitativos que os pesquisadores na área de CG podem usar ao abordar questões de pesquisa (Figura 1). A tipologia proposta envolve duas dimensões com base nas escolhas de desenho da pesquisa associadas a cada critério.

Na **primeira dimensão**, os pesquisadores decidem se o desenho de pesquisa inclui um grupo de controle. Se afirmativo (Figura 1, lado esquerdo), a escolha leva ao uso de métodos empíricos experimentais. Caso contrário, métodos empíricos não experimentais são escolhidos (Figura 1, lado direito). Na **segunda dimensão**, os pesquisadores decidem se usarão amostras mais ou menos representativas. Tanto para estudos experimentais como não experimentais, o pesquisador pode selecionar um método empírico que tenha menor ou maior representatividade da amostra. Em seguida, discutimos sobre cada método empírico alternativo.

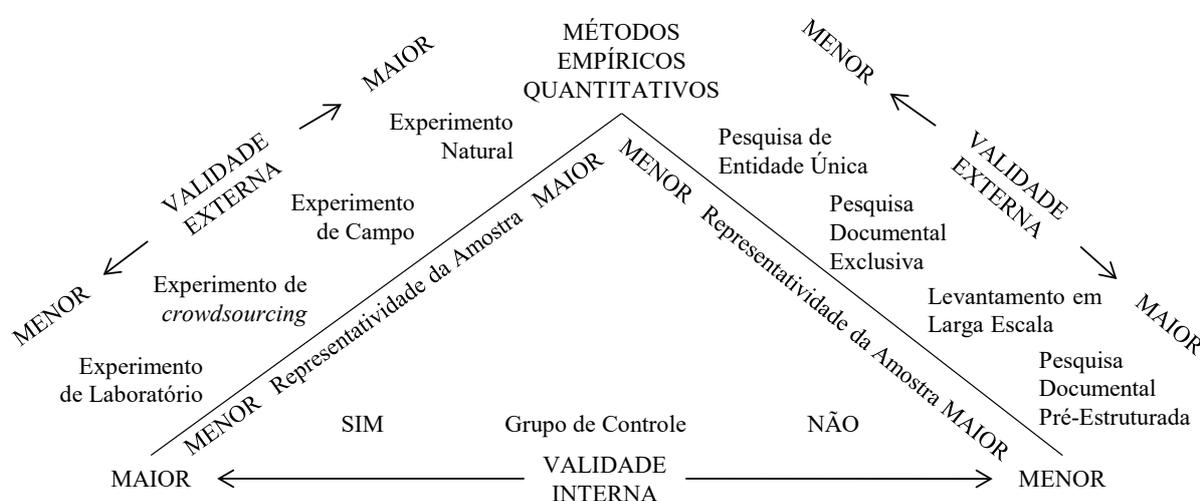


Figura 1. Tipologia proposta para pesquisa quantitativa em CG

3. Métodos Empíricos Quantitativos Alternativos

3.1 Experimento de Laboratório

Experimentos de laboratório têm como principais características a presença de grupo de controle, atribuição aleatória e o fato do tratamento (por exemplo, variável independente) ser manipulado (Bloomfield et al., 2016; Sprinkle & Williamson, 2006; Swieringa & Weick, 1982). Os participantes de experimentos de laboratório geralmente têm características semelhantes – mesmo curso de graduação, aproximadamente a mesma idade e experiência semelhante (Shadish et al., 2002; Sprinkle & Williamson, 2006). Os experimentos de laboratório podem ser executados em laboratórios ou, no caso de cientistas sociais, em salas onde o experimentador pode garantir o controle físico sobre os participantes, usando salas com espaços individuais para impedir que os participantes olhem uns para os outros ou adotando aparelhos tecnológicos durante o experimento.

O fato dos experimentos de laboratório usarem grupo de controle, atribuição aleatória e variáveis manipuladas melhora a validade interna desses estudos. Como já mencionado, os grupos de controle criam um contrafactual para eliminar explicações alternativas (Oehlert, 2003). Atribuição aleatória significa que dois ou mais grupos de unidades são, na média, probabilisticamente semelhantes entre si, de modo que quaisquer diferenças nos resultados entre os grupos experimental e de controle são provavelmente devidas ao tratamento e não a diferenças pré-existentes (Shadish et al., 2002). Variáveis manipuladas evidenciam que a causa propositalmente precede o efeito. Ao usar atribuição aleatória e variáveis manipuladas, o pesquisador pode controlar o cenário da pesquisa e também reduzir explicações alternativas, isolando os efeitos de variáveis de confusão (Sprinkle & Williamson, 2006).

No entanto, o ambiente controlado de experimentos de laboratório contribui para menor validade externa. A representatividade da amostra em experimentos de laboratório tende a ser bastante baixa porque normalmente são recrutados participantes homogêneos (por ex.: estudantes de graduação) usando amostragem não probabilística e intencional (Carpenter, Harrison e List, 2005). Na verdade, é desejável que os participantes de experimentos de laboratório sejam uma *tábula rasa* para que a única influência que eles experimentem seja o tratamento, o que os torna alunos adequados, principalmente se o estudo exigir conhecimentos específicos (por ex.: análise de demonstrações financeiras), mas não exigir experiência específica (Mortensen, Fisher, & Wines, 2012; Trottier & Gordon, 2018). Desta forma, os experimentos de laboratório são altamente criticados por sua artificialidade (Harrison, 2005; Carpenter et al., 2005).

Um exemplo de experimento de laboratório é o estudo de Haesebrouck (2021), que foca no processo de tomada de decisão de gerentes sobre os efeitos do esforço de aquisição de informações e propriedade psicológica induzida nos relatórios dos gerentes. Como parte de seu trabalho, os gerentes podem realizar um grande esforço para adquirir, sintetizar e analisar dados de várias fontes ou adquirir essas informações facilmente se a empresa possuir bons sistemas de compartilhamento de informações. Para testar as previsões, o autor usou um desenho experimental, no qual a aquisição de informação (fornecida vs. adquirida) e a saliência da honestidade no contexto do relatório (honestidade menos saliente vs. mais saliente) são manipuladas. O artigo garante validade interna por seguir vários procedimentos experimentais (ou seja, randomização, dados demográficos e observação direta dos participantes). O autor afirma que o experimento foca em testes de teoria e os resultados podem não ser generalizáveis para outros contextos.

Em geral, devido à presença de grupo de controle, atribuição aleatória e variáveis manipuladas, os experimentos de laboratório são considerados o método empírico com maior validade interna (Trochim et al., 2016) e com maior chance de estabelecer relações causais. Por outro lado, o uso de amostras menos representativas e ambientes artificiais torna os experimentos de laboratório o método empírico com a menor validade externa (Asay, Guggenmos, Kadous, Koonce, & Libby, 2021) e menor possibilidade de fazer afirmações sobre a generalização dos resultados.

3.2 Experimentos de *Crowdsourcing* (ou experimentos on-line)

Os experimentos de *crowdsourcing* são relativamente semelhantes aos experimentos de laboratório porque também incluem grupo de controle, atribuição aleatória e variável manipulada. Entretanto, os dois desenhos experimentais diferem quanto à localização e participantes. Os experimentos de *crowdsourcing* usam plataformas on-line (por ex.: MTurk, Prolific e CrowdFlower) em vez de laboratórios ou salas. O uso de tais plataformas permite que os pesquisadores recrutem participantes de grupos maiores, para além de estudantes de localizações geográficas limitadas (Peer, Brandimarte, Samat, & Acquisti, 2017).

Como os experimentos de *crowdsourcing* usam grupo de controle, atribuição aleatória e variáveis manipuladas, a validade interna desses estudos é aprimorada. No entanto, o uso de plataformas on-line impõe ameaças à validade interna porque os pesquisadores têm menor controle sobre o experimento, ou seja, menos controle sobre ruídos e vieses, como, por exemplo, falta de atenção ou esforço, ou mesmo a inclusão de participantes não qualificados (Bentley, 2021). Além disso, os pesquisadores podem enfrentar comportamentos fraudulentos por parte dos participantes, o que afeta a integridade e a confiabilidade dos dados (Aguinis & Ramani, 2021).

No entanto, o uso de um grupo mais heterogêneo de participantes contribui para aumentar a validade externa do *crowdsourcing* em relação aos experimentos de laboratório, devido a amostras mais representativas. Por um lado, as plataformas on-line permitem que os pesquisadores filtrem várias características demográficas que podem ajudar a restringir os participantes. Por outro lado, os pesquisadores podem examinar teorias recrutando participantes de diferentes idades, culturas e origens. Além disso, apesar do crescente uso de plataformas on-line, ainda há críticas sobre a confiabilidade desses dados em comparação com outros tipos de dados experimentais (Chmielewski & Kucker, 2020).

Um exemplo de experimento de *crowdsourcing* é o estudo de Murphy, Wynes, Hahn e Devine (2019), que investigam a motivação interna e externa para realizar relatórios honestos. Os autores usam os participantes do MTurk em um experimento em que eles têm a oportunidade e o incentivo para realizar relatórios não fidedignos, a fim de testar diferentes motivações por trás da honestidade. Como o objetivo é testar a motivação por trás de decisões, três manipulações são usadas: linha de base (grupo de controle) vs. recompensa vs. punição. Os autores explicam que os participantes do MTurk são adequados para os experimentos devido à natureza das tarefas (ou seja, não especializadas). Consistente com experimentos de *crowdsourcing*, o estudo combina validade interna e externa. Os autores asseguram a validade interna aplicando diferentes manipulações e a validade externa usando participantes com características diferentes.

Por um lado, os pesquisadores renunciam à parte da capacidade dos experimentos de laboratório, que é monitorar os participantes pessoalmente, o que torna o experimento de *crowdsourcing* de menor validade interna em comparação com os experimentos de laboratório. No entanto, o uso de plataformas on-line permite ao pesquisador alcançar um grupo maior e mais diversificado de participantes, o que melhora a validade externa dos experimentos de *crowdsourcing* em relação aos experimentos de laboratório.

3.3 Experimentos de Campo

Os experimentos de campo são estudos que utilizam o método experimental em que os tratamentos ou efeitos são observados por períodos mais longos (Lourenço, 2019; Bloomfield et al., 2016). Os experimentos de campo são semelhantes a outros experimentos, incluindo grupo de controle, atribuição aleatória e variável manipulada e a principal diferença é que eles são realizados em campo (ou seja, nas organizações) e os participantes são profissionais que trabalham em ambientes que ocorrem naturalmente e geralmente não estão cientes que fazem parte de um experimento (Lourenço, 2019; Floyd & List, 2016). Além disso, os experimentos de campo permitem que os pesquisadores usem dados proprietários e mensurem as variáveis para atingir o objetivo do estudo (Asay et al., 2021).

Como os experimentos de campo também adotam grupo de controle, atribuição aleatória e variáveis manipuladas, a validade interna desses estudos é aprimorada. No entanto, os experimentos de campo enfrentam maiores ameaças à validade interna devido a um menor nível de controle. A razão é que os pesquisadores não podem criar um ambiente artificial e isolado no qual os participantes fazem julgamentos e/ou decisões e as variáveis são manipuladas da mesma forma que em experimentos de laboratório (Lourenço, 2019).

O uso de ambientes naturais nos experimentos de campo, ou seja, ambientes que não são criados artificialmente pelo pesquisador em que os participantes não têm ciência de que fazem parte de um experimento e continuam realizando normalmente suas tarefas diárias, contribuem para aumentar a validade externa em relação aos experimentos de laboratório e *crowdsourcing*. No entanto, embora esses recursos possam aumentar a validade externa em termos de realismo, os experimentos de campo também são limitados a amostras menos representativas (por ex.: uma unidade de negócios, uma única empresa).

Um exemplo de experimento de campo é o estudo de Cronin, Erkens, Schloetzer e Tinsley (2021) sobre os efeitos do controle das percepções de falha no desempenho. Os autores manipulam uma mensagem em vídeo transmitida a vendedores durante reunião semanal em uma das 20 distribuidoras brasileiras de uma organização multinacional de vendas diretas. Na condição de tratamento, os vendedores assistem uma mensagem em vídeo do chefe regional incentivando-os a encarar o fracasso como uma “parte natural da história”. Na condição de controle os vendedores assistem o mesmo chefe regional resumindo a história da organização. Os autores conduziram o experimento por um período de quatro semanas, com grupos de controle e tratamento, usando dados proprietários (por ex.: comissão semanal de vendas). Embora os experimentos de campo tenham resultados mais generalizáveis em comparação aos experimentos de laboratório e *crowdsourcing*, os autores relatam as limitações do uso de métodos experimentais e os fatores de confusão associados aos estudos de campo.

Em geral, os experimentos de campo fornecem uma combinação potente: os benefícios da validade interna, consistentes com o método experimental, e dados do mundo real extraídos do campo (Bloomfield et al., 2016; Lourenço, 2019). Essa combinação oferece uma ótima mistura entre “controle e realismo, geralmente não alcançados em laboratório ou com dados não controlados” (Floyd & List, 2016, p. 438). Os experimentos de campo oferecem um nível mais baixo de validade interna devido ao controle reduzido associado a ambientes menos artificiais e isolados, mas esses custos vêm com o benefício de maior generalização para contextos que ocorrem naturalmente.

3.4 Experimento Natural

Experimentos naturais são eventos que ocorrem naturalmente (por ex.: choques exógenos) que não são manipulados, mas podem estabelecer um “contraste entre um tratamento e uma condição de comparação” (Shadish et al., 2002, p. 17). Os eventos que não acontecem por meio de intervenção natural (por ex.: enchentes, furacões) também podem ser tratados como experimentos naturais (por ex.: adoção de lei) (Mcvay, 2011). Portanto, experimentos naturais são viáveis quando os eventos ocorrem devido à intervenção humana ou natural e os pesquisadores podem comparar os resultados ex-ante com ex-post. Embora os experimentos naturais possam incluir um grupo de controle, aqueles não incluem atribuição aleatória ou variáveis manipuladas porque os eventos ocorrem naturalmente. Por essa razão, os experimentos naturais são considerados quasi-experimentos (Lourenço, 2019; Aguinis & Bradley, 2014). Além disso, os experimentos naturais podem se beneficiar da grande disponibilidade de dados pré-estruturados ou proprietários.

Embora os experimentos naturais também usem grupos de controle como contrafactuais para aumentar a validade interna, esse método de pesquisa tende a impor mais ameaças à validade interna. A principal razão é que os grupos de controle e tratamento não são projetados pelos pesquisadores, mas definidos de forma exógena. Portanto, os participantes não são designados aleatoriamente para as condições experimentais e, conseqüentemente, os pesquisadores não podem assumir que os grupos experimentais são equivalentes (Trochim et al., 2016). Assim, os pesquisadores que realizam experimentos naturais precisam encontrar métodos alternativos para minimizar ameaças à validade interna, como, por exemplo, usando métodos estatísticos sofisticados como Diferenças em Diferenças, Desenho de Regressão Descontínua e Controle Sintético (Lonati et al., 2018).

A extensão em que os resultados de experimentos naturais podem ser generalizados depende da unidade de observação. Quando a unidade de observação está no nível organizacional (caso mais comum), ou seja, quando o evento exógeno afetou uma grande amostra de organizações, a capacidade de obter amostras representativas e generalizar os resultados é maior. Entretanto, quando a unidade de observação está no nível de subunidade ou individual, o desafio com relação à validade externa é maior, pois os pesquisadores terão dificuldade em convencer seus leitores de que a organização estudada é de alguma forma semelhante a outras organizações.

Um exemplo de experimento natural em CG é o estudo de Flammer e Kacperczyk (2016), que examina os efeitos da orientação para inovação de partes interessadas (*stakeholders*) nas decisões de negócios da empresa. Os autores exploram a promulgação de estatutos constitucionais em nível estadual, particularmente a mudança estatutária da responsabilidade dos acionistas para as partes interessadas. De 1980 a 2006, 34 estados americanos adotaram estatutos constitucionais que se concentram na criação de valor para as partes interessadas. Os autores usam dados sobre a criação de patentes do National Bureau of Economic Research (NBER) Patent Data Project de 1976 a 2006 para avaliar a variável dependente, produtividade inovadora, medida como o número de patentes e citações dividido pelo número de funcionários da empresa. Os autores coletaram diversas variáveis de controle que poderiam atuar como fatores de confusão para lidar com as ameaças à validade interna. Esse estudo é consistente com experimentos naturais que fornecem maior validade externa devido ao uso de um grande conjunto de dados pré-estruturados disponíveis.

Em geral, semelhante aos experimentos de campo, os experimentos naturais combinam os benefícios da validade interna, incluindo um grupo de controle, e da validade externa, examinando eventos que ocorrem naturalmente (Bloomfield et al., 2016; Lourenço, 2019). Com isso, esse método empírico fornece graus intermediários de validade interna e validade externa (Roe & Just, 2009) que permitem afirmações equilibradas de relação causal, bem como a generalização dos resultados.

3.5 Pesquisa de Entidade Única

As pesquisas de entidade única são desenvolvidas com o uso de questionários submetidos aos respondentes de um ambiente organizacional. Os estudos que empregam esse método não incluem um grupo de controle, atribuição aleatória ou variáveis manipuladas. Nas pesquisas de entidade única, as informações qualitativas do contexto organizacional (ou seja, entrevistas) têm papel central na formação e na “calibração” dos instrumentos de pesquisa. Além disso, pesquisas de entidade única são normalmente administradas como transversal e o instrumento de pesquisa pode ser projetado com perguntas que capturam fatos em vez de percepções e opiniões. Esse método empírico tem mais chances de usar desenhos longitudinais e de limitar melhor os respondentes-alvo e usar amostras aleatórias. Ademais, o apoio da empresa pode influenciar a motivação dos entrevistados e minimizar possíveis vieses de resposta e não resposta. Os entrevistados em pesquisas de entidade única geralmente incluem funcionários (nível individual), grupos (nível de equipe) ou gerentes de subunidade (nível de subunidade).

O fato de que as pesquisas de entidade única não incluem um grupo de controle, atribuição aleatória ou variáveis manipuladas ameaçam a validade interna, considerando a separação não temporal de causa e efeito, bem como questões relacionadas à medição e ao desenho de pesquisa, como viés de método comum (Speklé & Widener, 2018). Em particular, a autosseleção é uma ameaça relevante à validade interna, seja pela empresa ou pelos respondentes que aceitam participar da pesquisa. O acesso ao campo, no entanto, pode reduzir as ameaças à validade interna de diferentes maneiras. Por exemplo, a compreensão do cenário antes da implementação da pesquisa permite que os pesquisadores obtenham respostas mais qualificadas e menos tendenciosas, identificando os possíveis respondentes mais bem informados e promovendo o seu engajamento para responder à pesquisa. Além disso, o uso de dados qualitativos pode ajudar a calibrar o instrumento de pesquisa e permitir escolhas adequadas de cargos/posições ideais para o estudo. Note também que a validade interna melhora substancialmente quando o desenho longitudinal é possível.

As pesquisas de entidade única enfrentam desafios ao fazer inferências para outras organizações, uma vez que o modelo de pesquisa depende do contexto. Embora os pesquisadores realizando pesquisas de entidade única geralmente obtenham taxas de resposta maiores do que em levantamentos em larga escala (Hiebl & Richter, 2018) e sejam mais capazes de abordar questões relacionadas ao viés de não resposta, os resultados não são generalizáveis para outras amostras devido às particularidades do contexto. Isso significa que os achados são apenas empiricamente aplicáveis em organizações com características e fenômenos muito semelhantes. Assim, considerando essas limitações, ao discutir a validade externa para pesquisas de entidade única, os pesquisadores geralmente estão se referindo à generalização da amostra para a população de indivíduos dentro desse ambiente organizacional em determinado período de tempo (por ex.: gerentes intermediários atuais ou trabalhadores de montagem) ou a generalização em nível teórico.

Um exemplo de pesquisa de entidade única em CG é Mucci, Frezatti e Bido (2021), que investiga a associação entre quatro características facilitadoras do projeto de orçamentos e percepções de utilidade dos gestores de uma organização que atua no setor de concessionárias de energia elétrica. A pesquisa de entidade única foi desenvolvida com uma amostra de 75 gerentes de nível médio de diferentes áreas de negócios (ou seja, finanças, operações e *marketing*) e operacionalizada com o apoio do gerente de orçamento da empresa. Os pesquisadores obtiveram uma taxa de resposta alta (42%) e seguiram vários procedimentos para mitigar as ameaças à validade interna (ou seja, modelo teórico rigoroso, variáveis de controle e viés de método comum).

Em geral, as pesquisas de entidade única enfrentam várias ameaças à validade interna, mas são capazes de lidar com parte delas definindo de forma estrita o modelo teórico e empírico à luz do contexto que está sendo investigado (Luft & Shields, 2002), bem como controlando parcialmente os efeitos de confusão que emergem do contexto. Por não terem um grupo de controle, as pesquisas de entidade única apresentam menor validade interna do que os métodos experimentais. Da mesma forma, um desenho limitado ao ambiente organizacional apresenta ameaças à validade externa. De fato, esse método empírico pode ser o de menor validade externa entre os métodos não experimentais. Dessa forma, enquanto pesquisas de entidade única têm um grau equilibrado de validade interna e externa, o grau geral de validade é menor do que o de experimentos de campo e experimentos naturais que também têm um grau de validade mais equilibrado.

3.6 Pesquisa Documental Exclusiva

A pesquisa documental exclusiva usa dados proprietários confidenciais de pesquisas de terceiros ou dados internos de empresas que podem ser acessados somente se o proprietário dos dados conceder acesso (Moers, 2006). Como um método não experimental, a pesquisa documental exclusiva não inclui grupo de controle, atribuição aleatória ou variáveis manipuladas. Se o conjunto de dados for obtido de terceiros, esse método empírico pode incluir um conjunto grande e heterogêneo de observações que podem proporcionar maior representatividade da amostra, o que aproxima a pesquisa documental exclusiva dos estudos documentais pré-estruturados. Por sua vez, se forem utilizados dados internos de uma empresa, esse método empírico está mais próximo de pesquisas de entidade única devido ao potencial uso de informações específicas do contexto que podem ajudar os pesquisadores na escolha de *proxies* apropriadas para as variáveis relevantes.

Como método não experimental, a pesquisa documental exclusiva enfrenta várias ameaças à validade interna associadas principalmente a autosseleção e endogeneidade (Lourenço, 2019). Além disso, a falta de um grupo de controle dificulta o descarte de explicações alternativas para os resultados. A pesquisa documental exclusiva geralmente segue procedimentos econométricos para lidar com ameaças à validade interna, como o uso de variáveis instrumentais. Essas ameaças são particularmente graves quando pesquisas de terceiros são usadas. Quando os dados internos da empresa são utilizados, os pesquisadores podem coletar dados não estruturados adicionais e estruturá-los para criar medidas adequadas para abordar questões de pesquisa relevantes para um determinado estudo, o que dá mais flexibilidade na busca de *proxies* adequadas para as variáveis de interesse. Além disso, os dados proprietários podem ser combinados com entrevistas de campo para ajudar a identificação de *proxies* adequados para variáveis relevantes.

A pesquisa documental exclusiva envolve amostras grandes e abrangentes (Moers, 2006). No entanto, sua capacidade de fazer inferências baseadas em amostras observadas para amostras localizadas em outros lugares ou em outros momentos, pode ser mais difícil, particularmente em relação aos dados internos de uma empresa. A representatividade da amostra, assim como a capacidade de generalizar os resultados de amostras observadas para outras amostras, é um desafio importante quando são usados os dados internos de uma empresa. A razão é que as observações de análise em nível de subunidade ou individual (por ex.: funcionário) podem ser exclusivas das características da empresa estudada, dificultando a generalização dos resultados para empresas com características diferentes. Por essa razão, semelhante aos experimentos de laboratório, a pesquisa documental exclusiva pode melhor argumentar a favor da generalização para a teoria que está sendo testada.

Ikäheimo, Kallunki, University e Schiehl (2018) é uma pesquisa documental exclusiva sem grupo de controle, usando dados internos de uma empresa. Ikaheimo et al. (2018) examinam a relação entre incentivos baseados em desempenho para trabalhadores profissionais e a lucratividade futura da empresa e se essa relação depende da complexidade da tarefa. Eles usam um grande conjunto de dados proprietários de remuneração de uma pesquisa administrada pela Confederação das Indústrias Finlandesas. O conjunto de dados inclui mais de 564.000 observações individuais de funcionários por ano e 7.820 observações por ano da empresa durante os anos de 2002 a 2011. Os autores realizam verificações de robustez para lidar com questões de endogeneidade e, assim, aumentar a validade interna. Considerando um grande conjunto de dados, há alta capacidade de generalização dos resultados para outras organizações que usam esquemas de incentivos e planejam alterá-los, ou ainda não possuem esquemas de incentivos, mas planejam adotá-los.

Por um lado, as ameaças à validade interna são maiores quando as pesquisas documentais exclusivas incluem pesquisas de terceiros, devido ao viés de autoseleção e problemas de endogeneidade. Contudo, a representatividade da amostra e, portanto, a validade externa, é maior. Como veremos, essas características aproximam esse método empírico de pesquisas documentais pré-estruturadas. Por outro lado, quando os dados internos de uma empresa são usados, as preocupações de validade interna são sanadas pelo uso de informações específicas do contexto, mas a validade externa é prejudicada, pois o conjunto de dados é específico de uma única organização. Nesse caso, esses recursos aproximam esse método empírico das pesquisas de entidade única.

3.7 Levantamento em larga escala

Os levantamentos em larga escala são normalmente usados em pesquisas quantitativas de CG (Van der Stede et al., 2006; Speklé & Widener, 2018). Nesse método os dados são coletados usando questionários submetidos a um amplo conjunto de respondentes em potencial. Ainda assim, costuma ser um desafio para os pesquisadores de CG a obtenção de um grande número de respostas. A representatividade da amostra depende de seleção aleatória e taxa de resposta; entretanto, a maioria dos estudos na área de CG utiliza amostras de conveniência. Semelhante a outros métodos não experimentais, os levantamentos em larga escala não incluem grupo de controle, atribuição aleatória ou variáveis manipuladas. Os levantamentos em larga escala geralmente são projetados de forma transversal e as variáveis relevantes são eliciadas no instrumento de pesquisa (Bloomfield et al., 2016). Mesmo quando envolvem dados longitudinais, os levantamentos em larga escala podem sofrer com a falta de respostas, pois os entrevistados podem não estar disponíveis para participar da segunda ou terceira rodada da pesquisa. Por fim, os levantamentos em larga escala geralmente consistem de dados que expressam fatos, opiniões ou percepções, considerando diferentes níveis de análise, geralmente nos níveis organizacional ou de unidade de negócios.

Como um método não experimental, levantamentos em larga escala enfrentam várias ameaças à validade interna, principalmente associadas ao viés de autoseleção e problemas de endogeneidade. Não há contrafactual ou diferença de tempo entre causa e efeito. Dessa forma, é mais viável argumentar a favor da diferença de tempo entre causa e efeito quando dados longitudinais são coletados, mas mortalidade é uma ameaça relevante. Para lidar com ameaças à validade interna, os pesquisadores que realizam levantamentos em larga escala podem adotar várias soluções, como definir uma população teórica, população-alvo e respondentes-alvo, usar diferentes tipos de respostas (fatos, opiniões, etc.), assim como desenhar o instrumento de pesquisa cuidadosamente. No entanto, embora eficazes, essas soluções não superam completamente as ameaças à validade interna.

Os levantamentos em grande escala se beneficiam do uso de grandes amostras e, portanto, fornecem evidências que podem ser generalizadas para outras amostras. No entanto, os pesquisadores geralmente estão cientes de que o potencial de generalização dos resultados depende da representatividade da amostra, viés de não resposta e taxas de resposta (Hiebl & Richter, 2018; Spekle & Widener, 2018). Além disso, a capacidade de levantamentos em larga escala para generalizar resultados depende da amostra ser probabilística ou não; a segunda estratégia é a mais comum. Dessa forma, as ferramentas disponíveis para aumentar validade externa são diferentes daquelas aplicadas para outros métodos.

Um exemplo de levantamento em larga escala em CG é Bedford, Spekle e Widener (2022) em que um levantamento em larga escala é conduzido com gerentes de Unidades de Negócios (UN) na Holanda, com uma amostra final de 83 entrevistados. Eles estudaram como as empresas alteram o aperto orçamentário em resposta à crise global e as implicações desse aperto para estresse e exaustão emocional entre os funcionários, considerando também um projeto de orçamento facilitador como moderador para essa relação. Bedford et al. (2022) realizaram um estudo transversal usando questionários on-line, dirigidos a uma população de 172 gerentes de UN. Embora o tamanho da amostra seja relativamente pequeno, eles obtiveram uma alta taxa de resposta (48,3%). Os autores seguem vários procedimentos para mitigar ameaças à validade interna e à validade externa e abordam viés de não resposta e viés de método comum, além de usar instrumentos validados e variáveis de controle.

Em geral, esse método empírico enfrenta maiores ameaças à validade interna do que todos os métodos empíricos anteriores, sejam experimentais ou não. Em comparação com a pesquisa de entidade única e pesquisa documental exclusiva, os levantamentos em larga escala não consideram dados contextuais (ou seja, particularidades organizacionais), gerando problemas relacionados à definição de respondentes adequados, adequação dos instrumentos de pesquisa e viés associado a respostas individuais (por ex.: efeito halo, deseabilidade social, falta de conhecimento). Por outro lado, em comparação com os outros métodos não experimentais discutidos até agora, os resultados de pesquisas em larga escala têm maior nível de validade externa devido à participação de um grande conjunto de respondentes.

3.8 Pesquisa Documental Pré-Estruturada

A pesquisa documental pré-estruturada adota dados arquivísticos que são registrados e estruturados por terceiros, cujo objetivo principal não é a pesquisa acadêmica (Bloomfield et al., 2016; Moers, 2006). Pesquisas documentais pré-estruturadas não incluem grupo de controle, amostragem aleatória ou variáveis manipuladas. As variáveis relevantes são operacionalizadas através do uso de *proxies* definidas a partir do conjunto de dados disponível. O conjunto de dados normalmente inclui observações referentes a vários respondentes, principalmente em nível organizacional. Esses estudos são o principal método de pesquisa utilizado na literatura contábil (Bloomfield et al., 2016). Para estudos de CG, a falta de dados públicos disponíveis torna mais difícil o uso de estudos de documentos pré-estruturados comparados aos métodos de pesquisa alternativos (Hesford et al., 2006; Moers, 2006; Aguiar, 2018).

Como método não experimental, esse método enfrenta uma dificuldade para descartar explicações alternativas porque as associações observadas entre as variáveis de interesse podem ser atribuídas à causalidade reversa, variáveis correlacionadas omitidas ou forma funcional incorretamente especificada (Gassen, 2014). Semelhante à pesquisa documental usando dados proprietários, a pesquisa usando dados arquivísticos pré-estruturados enfrenta várias ameaças à validade interna associadas ao viés de seleção e endogeneidade que os pesquisadores tentam resolver usando ferramentas econométricas (por ex.: variáveis instrumentais) (Lourenço, 2019).

Considerando que a pesquisa documental pré-estruturada usa grandes amostras (Das et al., 2016; Moers, 2006), sua capacidade de fazer inferências baseadas nas amostras observadas para amostras localizadas em outros lugares e em outros momentos é relativamente alta. A unidade de observação tende a ser altamente representativa, de modo que os resultados podem ser generalizados para a população de interesse. No entanto, o viés de seleção pode representar desafios para a representatividade da amostra nesses estudos, em particular, viés de autoseleção, uma vez que a divulgação de informações de CG não é aleatória (Moers, 2006).

Laviers, Sandvik e Xu, (2021) é um exemplo de pesquisa documental pré-estruturada, pois não inclui grupo de controle e usa vários grandes conjuntos de dados disponíveis. Laviers et al. (2021) examinam as reações dos investidores às divulgações voluntárias do índice de remuneração do CEO. Eles coletaram declarações *proxy* de empresas listadas no índice Standard & Poor's 1500 com divulgação obrigatória do índice de remuneração do CEO e classificaram as empresas como tendo um índice de remuneração baixo, médio ou alto. Os autores combinaram as informações da taxa de remuneração do CEO com informações coletadas de plataformas de banco de dados, como retornos de ações do CRSP, informações financeiras do Compustat e remuneração dos executivos da Execucomp. Dada a falta de grupo de controle, Laviers et al. (2021) realizaram várias análises adicionais e testes de robustez usando diferentes especificações empíricas para aumentar a validade interna. Em relação à generalização dos resultados, os autores tratam da autoseleção utilizando um procedimento de estimação pelo qual incluem a razão de Mills inversa (Heckman, 1979).

Por um lado, devido à falta de grupo de controle, de atribuição aleatória e de variáveis manipuladas, os estudos documentais pré-estruturados enfrentam várias ameaças à validade interna que são tratadas com o uso de ferramentas econométricas. De fato, sugerimos que os esse é o método empírico com menor validade interna. Por outro lado, devido ao uso de um grande conjunto de dados, incluindo observações de organizações de diferentes setores, idades, tamanhos, etc., as pesquisas documentais pré-estruturadas tendem a incluir amostras altamente representativas, aumentando sua capacidade de fazer afirmações sobre generalizações.

3.9 Sumário

Resumimos, nesta seção, a tipologia proposta de métodos empíricos quantitativos, enfatizando as implicações em termos das vantagens e desvantagens (*trade-offs*) entre validade interna e validade externa associadas a cada método. A Figura 2 apresenta cada método empírico posicionado de acordo com os benefícios relativos da validade interna vs. validade externa.

A Figura 2 mostra que, em geral, os estudos experimentais apresentam maior validade interna devido à presença de grupo de controle. Em particular, os experimentos de laboratório são o método com maior validade interna e menor validade externa. Passando de laboratório para experimentos naturais, há um aumento na validade externa devido ao uso de amostras mais representativas (experimentos naturais), ambientes menos artificiais (experimentos de campo) e participantes mais heterogêneos (experimentos de *crowdsourcing*), associada a uma diminuição da validade causada pelo uso de configurações menos controladas (experimentos naturais), ambientes naturais (experimentos de campo) e plataformas on-line (experimentos de *crowdsourcing*).

Note também que a pesquisa documental pré-estruturada tem a maior validade externa devido ao uso de grandes conjuntos de dados e a menor validade interna porque apresenta problemas de seleção e endogeneidade. Consideramos que a pesquisa documental pré-estruturada e levantamentos em larga escala têm graus de validade interna semelhantes, enquanto a validade externa é menor devido a baixas taxas de resposta para levantamentos em larga escala. A pesquisa documental com dados proprietários e as pesquisas de entidade única são os métodos não experimentais com a menor validade externa devido ao uso de um conjunto de dados específico para uma única organização e, ao mesmo tempo, com a maior validade interna devido à consideração de informação específica do contexto.

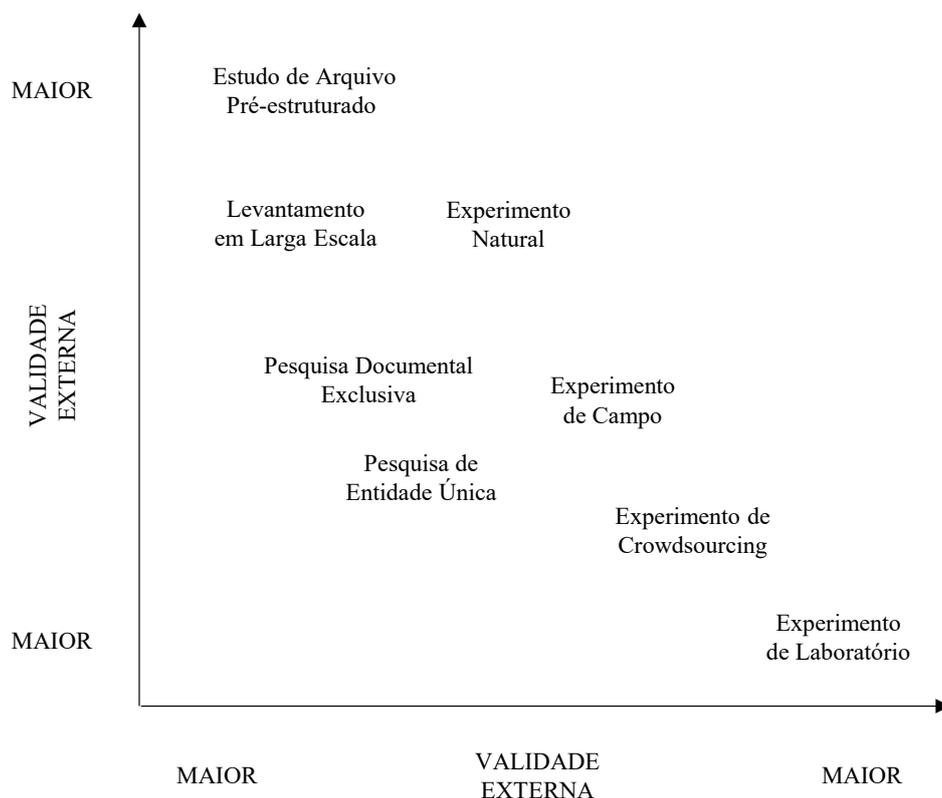


Figura 2. Vantagens e desvantagens entre a validade interna e validade externa para cada método empírico.

4. Conclusão

Este estudo propõe uma tipologia de pesquisa empírica quantitativa em CG e discute as principais implicações da tipologia proposta para a estrutura de validade. Especificamente, propomos oito métodos empíricos quantitativos baseados em dois critérios (grupo de controle e representatividade da amostra), que os pesquisadores de CG podem usar ao abordar questões de pesquisa, incluindo os métodos experimental (laboratório, *crowdsourcing*, experimentos campo e natural) e não experimental (pesquisas de entidade única, pesquisa documental exclusiva, levantamentos em grande escala e pesquisa documental pré-estruturada). Nos concentramos nas implicações da tipologia proposta para os *trade-offs* entre validade interna e validade externa.

A tipologia proposta e as implicações para validade podem beneficiar pesquisadores juniores da área de CG no Brasil de várias maneiras. Em primeiro lugar, este estudo pode ajudar a ampliar o escopo das questões de pesquisa a serem investigadas. Por exemplo, se o pesquisador estiver interessado em examinar o papel de declarações de valor no comportamento dos funcionários, métodos alternativos podem ser usados como, por exemplo, projetar experimentos de laboratório e manipular as declarações de valor para capturar as respostas dos funcionários; acessar organizações individuais para usar os dados disponíveis sobre a compreensão dos valores organizacionais pelos funcionários; ou usar levantamentos para coletar as respostas de funcionários localizados em diferentes organizações sobre suas percepções de valores organizacionais e respostas comportamentais (por ex.: comprometimento com metas).

Em segundo lugar, este estudo pode ajudar os pesquisadores de CG a construir programas de pesquisa envolvendo o uso sequencial de métodos alternativos para abordar a mesma questão de pesquisa, com o benefício evidente de replicar e generalizar os resultados. Por exemplo, os efeitos comportamentais de controles orçamentários rígidos examinados em levantamentos em larga escala podem ser examinados mais profundamente em projetos experimentais para oferecer evidências mais fortes de causalidade. Em terceiro lugar, este estudo pode ajudar os pesquisadores de CG a identificar métodos alternativos para examinar questões de pesquisa que não podem ser abordadas de forma adequada por métodos convencionais (por ex.: pesquisas em larga escala). Por exemplo, se a questão de pesquisa for examinar os efeitos do Covid-19 no uso de informações contábeis pelos funcionários para a tomada de decisões, os pesquisadores podem ter acesso a uma organização e coletar dados proprietários sobre a frequência de uso de informações contábeis antes e após o início da pandemia em um experimento natural e, em seguida, examinar se esse uso foi alterado.

Assim, chamamos a atenção de pesquisadores juniores na área de CG para as implicações do método empírico selecionado para os *trade-offs* entre as validades interna e externa. Ao escolher um método em detrimento de outro, o pesquisador também está selecionando a importância relativa atribuída à validade interna vs. validade externa. Os experimentos de laboratório aumentam a validade interna, enquanto impõem desafios à validade externa. Inversamente, a pesquisa documental pré-estruturada aumenta a validade externa, mas impõe maiores ameaças à validade interna. Portanto, enquanto afirmações causais são mais viáveis quando experimentos de laboratório são usados em vez de pesquisa documental pré-estruturada, visto que o pesquisador segue procedimentos apropriados para aumentar a validade interna, afirmações sobre generalizações são mais viáveis quando a pesquisa documental pré-estruturada é usada em vez de experimentos de laboratório, mais uma vez considerando que procedimentos adequados são adotados para aumentar a validade externa.

De qualquer forma, recomendamos que os pesquisadores juniores na área de CG se preocupem com suas escolhas de desenho de pesquisa, uma vez que “Os resultados da pesquisa são inevitavelmente o produto do desenho da pesquisa” (Dyckman & Zeff, 2014, p. 697). Em particular, dois pontos podem ajudar na escolha do método empírico. O primeiro é que a escolha do método empírico não deve ser motivada pela disponibilidade de um conjunto de dados específico ou pela curiosidade em usar um método específico, mas, sim, pela questão de pesquisa (Kinney 2019; Dyckman & Zeff, 2014). Embora explorar os diferentes métodos possa ser valioso para adquirir novas habilidades de pesquisa, a questão de pesquisa deve vir em primeiro lugar para orientar essa decisão. Os benefícios em termos de maiores chances de publicação são maiores quando os pesquisadores obtêm conhecimento profundo sobre como usar métodos empíricos específicos, pois cada método envolve um protocolo diferente a ser seguido para lidar com ameaças à validade. Assim, independentemente do método escolhido, espera-se que os pesquisadores apliquem protocolos de pesquisa adequados na realização do estudo, de acordo com as melhores práticas estabelecidas na área. Em segundo lugar, os pesquisadores estarão em melhor situação escolhendo um método empírico que seja viável dadas as restrições existentes, como disponibilidade de dados, acesso a organizações, tempo e recursos financeiros.

Os pesquisadores na área de CG estão se tornando cada vez mais criativos para coletar dados por meio de projeto de pesquisa monométodo ou multimétodo. Para estudos monométodos, os pesquisadores na área de CG estão tirando vantagem da validade interna associada a estudos experimentais e a validade externa associada às pesquisas documentais realizando estudos quasi-experimentais, usando dados proprietários e escolhendo desenhos típicos de estudos experimentais, com pré e pós-teste (por ex.: Brügger, Grabner e Sedatole, 2021; Forker, Grabner e Sedatole, 2020). Para estudos multimétodos, os pesquisadores de CG estão combinando diferentes procedimentos de coleta de dados (por ex.: Bol, Braga de Aguiar, & Lill, 2020; Wouters & Wilderom, 2008). O principal benefício de combinar diferentes métodos é aumentar as fontes de dados relevantes e fornecer resultados mais robustos usando métodos que se complementam, como dados proprietários sobre o desempenho dos funcionários, combinados com medidas perceptivas sobre a motivação dos funcionários, capturadas através de pesquisas de entidade única.

Enquanto discutimos as implicações de cada método para os *trade-offs* entre as validades interna e externa, reconhecemos que as validades do constructo e da conclusão também são essenciais para estabelecer resultados válidos. A validade de construto refere-se à “validade das inferências sobre os construtos de ordem superior que representam particularidades da amostragem”, enquanto a validade de conclusão refere-se à “validade das inferências sobre a correlação (covariação) entre tratamento e resultado” (Shadish et al., 2002, p.38). Em geral, esses quatro tipos de validade são interdependentes e o uso da estrutura de validade preditiva, também conhecida como Libby Boxes, pode ser útil para os pesquisadores visualizarem melhor o seu projeto de pesquisa, bem como identificar possíveis ameaças à validade.

Referências

- Aguiar, A. B. (2017). Pesquisa Experimental Em Contabilidade: Propósito, Desenho E Execução. *Advances in Scientific and Applied Accounting*, 10(2), 224–244. <https://doi.org/10.14392/asaa.2017100206>
- Aguiar, A. B. (2018). O pequeno mundo da pesquisa em contabilidade gerencial no Brasil: discussão sobre desenhos alternativos de pesquisa. *Revista de Contabilidade e Organizações*, 12, e151933. <https://doi.org/10.11606/issn.1982-6486.rco.2018.151933>
- 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>
- Aguinis, H., & Ramani, R. S. (2021). MTurk Research : Review and Recommendations. *Journal of Management*, 47(4), 823–837. <https://doi.org/10.1177/0149206320969787>
- Asay, H. S., Guggenmos, R. D., Kadous, K., Koonce, L., & Libby, R. (2021). Theory Testing and Process Evidence in Accounting Experiments. *The Accounting Review*.
- Bedford, D. S., Spekle, R. F., & Widener, S. K. (2022). Accounting , Organizations and Society Budgeting and employee stress in times of crisis : Evidence from the Covid-19 pandemic. *Accounting, Organizations and Society*, (xxxx). <https://doi.org/10.1016/j.aos.2022.101346>
- Bentley, J. W. (2021). Improving the Statistical Power and Reliability of Research Using Amazon Mechanical Turk. *Accounting Horizons*, 35(4), 45–62. <https://doi.org/10.2308/HORIZONS-18-052>
- Bloomfield, R., Nelson, M. W., & Soltes, E. (2016). Gathering Data for Archival, Field, Survey, and Experimental Accounting Research. *Journal of Accountig Research*, 54(2), 341–395. <https://doi.org/10.1111/1475-679X.12104>
- Bol, J. C., Braga de Aguiar, A., & Lill, J. B. (2020). Peer-Level Calibration of Performance Evaluation Ratings : Are There Winners or Losers ?

- Brüggen, A., Grabner, I., & Sedatole, K. L. (2021). The Folly of Forecasting: The Effects of a Disaggregated Demand Forecasting System on Forecast Error, Forecast Positive Bias, and Inventory Levels. *The Accounting Review*, 96(2), 127–152. <https://doi.org/10.2308/tar-2018-0559>
- Carpenter, J. P., Harrison, G. W., & List, J. A. (2005). Field experiments in economics: An introduction. In: Harrison, G. W., Carpenter, J., & List, J. A. (Eds.) *Field Experiments in Economics (Research in Experimental Economics, Vol. 10)*, Emerald Group Publishing Limited, Bingley, pp. 1-15. [https://doi.org/10.1016/S0193-2306\(04\)10001-X](https://doi.org/10.1016/S0193-2306(04)10001-X)
- Chmielewski, M., & Kucker, S. C. (2020). An MTurk Crisis? Shifts in Data Quality and the Impact on Study Results. *Social Psychological and Personality Science*, 11(4), 464–473. <https://doi.org/10.1177/1948550619875149>
- Cronin, M., Erkens, D. H., Schloetzer, J. D., & Tinsley, C. H. (2021). How controlling failure perceptions affects performance: Evidence from a field experiment. *Accounting Review*, 96(2), 205–230. <https://doi.org/10.2308/TAR-2018-0146>
- Das, R., Jain, K. K., & Mishra, S. K. (2016). Archival Research: A Neglected Method in Organization Studies. *Benchmarking: An International Journal*. <https://doi.org/https://doi.org/10.1108/BIJ-08-2016-0123>
- Dyckman, T. R., & Zeff, S. A. (2014). Some methodological deficiencies in empirical research articles in accounting. *Accounting Horizons*, 28(3), 695–712. <https://doi.org/10.2308/acch-50818>
- Flammer, C., & Kacperczyk, A. (2016). The impact of stakeholder orientation on innovation: Evidence from a natural experiment. *Management Science*, 62(7), 1982–2001. <https://doi.org/10.1287/mnsc.2015.2229>
- Floyd, E., & List, J. A. (2016). Using Field Experiments in Accounting and Finance. *Journal of Accounting Research*, 54(2), 437–475. <https://doi.org/10.1111/1475-679X.12113>
- Forker, E., Grabner, I., & Sedatole, K. (2020). *Does learning by disaggregating accelerate learning by doing? The effect of forecast disaggregation on the rate of improvement in demand forecast accuracy.*
- Frezatti, F., Aguiar, A. B., Wanderley, C. A., & Malagueño, R. (2015). A pesquisa em contabilidade gerencial no Brasil: desenvolvimento, dificuldades e oportunidades. *Revista Universo Contábil*, 11(1), 47–68. <http://dx.doi.org/10.4270/ruc.2015147-68>
- Gassen, J. (2014). Causal inference in empirical archival financial accounting research. *Accounting, Organizations and Society*, 39(7), 535–544. <https://doi.org/10.1016/j.aos.2013.10.004>
- Haesebrouck, K. (2021). The Effects of Information Acquisition Effort, Psychological Ownership, and Reporting Context on Opportunistic Managerial Reporting*. *Contemporary Accounting Research*, 38(4), 3085–3112. <https://doi.org/10.1111/1911-3846.12712>
- Harrison, G. W. (2005). Field experiments and control. In: Harrison, G. W., Carpenter, J., & List, J. A. (Eds.) *Field Experiments in Economics (Research in Experimental Economics, Vol. 10)*, Emerald Group Publishing Limited, Bingley, pp. 17-50. [https://doi.org/10.1016/S0193-2306\(04\)10002-1](https://doi.org/10.1016/S0193-2306(04)10002-1)
- Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153–161.
- Hesford, J. W., Lee, S. H., Van der Stede, W. A., & Young, S. M. (2006). Management Accounting: A Bibliographic Study. *Handbooks of Management Accounting Research*, 1, 3–26. [https://doi.org/10.1016/S1751-3243\(06\)01001-7](https://doi.org/10.1016/S1751-3243(06)01001-7)
- Hiebl, M. R. W., & Richter, J. F. (2018). Response Rates in Management Accounting Survey Research. *Journal of Management Accounting Research*, 30(2), 59–79. <https://doi.org/10.2308/jmar-52073>
- Ikäheimo, S., Kallunki, J.-P., University, S. M., & Schiehl, E. (2018). Do White-Collar Employee Incentives Improve Firm. *Journal of Management Accounting Research*, 30(3), 95–115. <https://doi.org/10.2308/jmar-51902>

- Kinney, W. R. (2019). The Kinney Three Paragraphs (and More) for Accounting Ph.D. Students. *Accounting Horizons*, 33(4), 1–14. <https://doi.org/10.2308/acch-52451>
- Laviers, L., Sandvik, J., & Xu, D. (2021). *CEO Pay Ratio Voluntary Disclosures and Investor Reactions*.
- Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management*, 64(April), 19–40. <https://doi.org/10.1016/j.jom.2018.10.003>
- Lourenço, S. M. (2019). Field Experiments in Managerial Accounting Research. *Foundations and Trends® in Accounting*, 14(1), 1–72. <https://doi.org/10.1561/14000000059>
- Luft, J., & Shields, M. (2002). Zimmerman's Contentious Conjectures: Describing the Present and Prescribing the Future of Empirical Management Accounting Research. *European Accounting Review*, 11(4), 795–803. <https://doi.org/10.1080/0963818022000047091>
- Luft, J., & Shields, M. D. (2014). Subjectivity in developing and validating causal explanations in positivist accounting research. *Accounting, Organizations and Society*, 39(7), 550–558. <https://doi.org/10.1016/j.aos.2013.09.001>
- Mcvay, S. E. (2011). Discussion of Do Control Effectiveness Disclosures Require SOX 404(b) Internal Control Audits? A Natural Experiment with Small U.S. Public Companies. *Journal of Accounting Research*, 49(2), 449–456. <https://doi.org/10.1111/j.1475-679X.2011.00403.x>
- Moers, F. (2006). Doing Archival Research in Management Accounting. *Handbooks of Management Accounting Research*, 1, 399–413. [https://doi.org/10.1016/S1751-3243\(06\)01016-9](https://doi.org/10.1016/S1751-3243(06)01016-9)
- Mortensen, T., Fisher, R., & Wines, G. (2012). Students as surrogates for practicing accountants: Further evidence. *Accounting Forum*, 36(4), 251–265. <https://doi.org/10.1016/j.acfor.2012.06.003>
- Mucci, D. M., Frezatti, F., & Bido, D. de S. (2021). Enabling design characteristics and budget usefulness. *RAUSP Management Journal*, 56, 38–54. <https://doi.org/10.1108/RAUSP-04-2019-0058>
- Murphy, P. R., Wynes, M., Hahn, T.-A., & Devine, P. G. (2019). Why are People Honest? Internal and External Motivations to Report Honestly. *Contemporary Accounting Research*, 53(9), 1689–1699. <https://doi.org/10.1017/CBO9781107415324.004>
- Nascimento, A. R., Junqueira, E., & Martins, G. A. (2010). Pesquisa Acadêmica em Contabilidade Gerencial no Brasil: Análise e Reflexões sobre Teorias, Metodologias e Paradigmas. *Revista de Administração Contemporânea*, 14(6), 1113–1133.
- Oehlert, G. W. (2003). A First Course in Design and Analysis of Experiments. In *The American Statistician* (Vol. 57). <https://doi.org/10.1198/tas.2003.s210>
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk : Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153–163. <https://doi.org/10.1016/j.jesp.2017.01.006>
- Roe, B. E., & Just, D. R. (2009). Internal and external validity in economics research: Tradeoffs between experiments, field experiments, natural experiments, and field data. *American Journal of Agricultural Economics*, 91(5), 1266–1271. <https://doi.org/10.1111/j.1467-8276.2009.01295>
- Saunders, M. N. K., Lewis, P., & Thornhill, A. (2019). Research Methods for Business Students. In *Pearson* (8th ed., Vol. 3). <https://doi.org/10.1108/qmr.2000.3.4.215.2>
- Shadish, W. R., Cook, T., & Campbell, D. (2002). Experimental and quasi-experimental designs for generalized causal inference. In *Houghton Mifflin Company*. <https://doi.org/10.1016/j.evalprogplan.2004.01.006>
- Smith, M. (2022). *Research Methods in Accounting* (6th ed.). SAGE.

- Spekle, R. F., & Widener, S. K. (2018). Challenging Issues in Survey Research: Discussion and Suggestions. *Journal of Management Accounting Research*, 30(2), 3–21. <https://doi.org/10.2308/jmar-51860>
- Sprinkle, G. B., & Williamson, M. G. (2006). Experimental research in managerial accounting. *Handbooks of Management Accounting Research*, 1, 415-444. [https://doi.org/10.1016/S1751-3243\(06\)01017-0](https://doi.org/10.1016/S1751-3243(06)01017-0)
- Swieringa, R. J., & Weick, K. E. (1982). An Assessment of Laboratory Experiments in Accounting. *Journal of Accounting Research*, 20, 56–101.
- Trochim, W. M., Donnelly, J. P., & Arora, K. (2016). *Research Methods - The essential knowledge base* (2nd ed.). Cengage Learning.
- Trottier, K., & Gordon, I. M. (2018). Students as surrogates for managers: Evidence from a replicated experiment. *Canadian Journal of Administrative Sciences*, 35(1), 146–161. <https://doi.org/10.1002/cjas.1377>
- Van der Stede, W. A., Young, S. M., & Chen, C. X. (2006). Doing Management Accounting Survey Research. *Handbooks of Management Accounting Research*, 1, 445–478. [https://doi.org/10.1016/S1751-3243\(06\)01018-2](https://doi.org/10.1016/S1751-3243(06)01018-2)
- Van der Stede, W. A. (2014). A manipulationist view of causality in cross-sectional survey research. *Accounting, Organizations and Society*, 39(7), 567-574. <https://doi.org/10.1016/j.aos.2013.12.001>
- Wouters, M., & Wilderom, C. (2008). Developing performance-measurement systems as enabling formalization: A longitudinal field study of a logistics department. *Accounting, Organizations and Society*, 33(4–5), 488–516. <https://doi.org/10.1016/j.aos.2007.05.002>