

## PESQUISA EXPERIMENTAL EM CONTABILIDADE: PROPÓSITO, DESENHO E EXECUÇÃO

Andson Braga de Aguiar<sup>1</sup>

**Resumo:** Este artigo discute aspectos centrais no desenho e na execução de experimentos, essenciais para o apropriado uso desse método de pesquisa em contabilidade. Para ilustrar os aspectos centrais, são usados exemplos representativos de pesquisas experimentais na área de contabilidade publicadas em periódicos internacionais. A contribuição deste estudo está em oferecer informações adicionais a pesquisadores interessados em aprimorar o conhecimento e experiência no uso de experimentos como método alternativo de pesquisa quantitativa em contabilidade no Brasil.

**Palavras-chave:** Experimento, Desenho, Execução, Contabilidade.

---

<sup>1</sup> e-mail: abraga@usp.br - Universidade de São Paulo

▪ DOI: <http://dx.doi.org/10.14392/asaa.2017100206>

▪ Artigo submetido em: 13/09/2016. Submetido a nova rodada em: 09/11/2016. Aceito em: 06/03/2017

## EXPERIMENTAL RESEARCH IN ACCOUNTING: PURPOSE, DESIGN AND EXECUTION

**Abstract** This paper discusses central features in the design and execution of experiments which are fundamental for the appropriate use of this research method in accounting. To illustrate the central features, I use representative examples of experimental research published in international accounting journals. This paper contributes to Brazilian researchers willing to acquire more knowledge and apply experimental designs as an alternative method of quantitative research in accounting.

**Keywords:** Experiment, Design, Execution, Accounting.

## 1. INTRODUÇÃO

Pesquisa experimental tem sido amplamente utilizada em diversas áreas do conhecimento. O amplo uso de experimentos está associado ao seu papel proeminente no desenvolvimento de pesquisas que buscam entender fatores que explicam mudanças de comportamento. No caso da contabilidade, apesar de ser relativamente mais recente e de receber mais ou menos atenção ao longo das últimas décadas (Hesford, Lee, Van der Stede e Young, 2007), o uso de pesquisa experimental tem cumprido papel importante na explicação de mudanças comportamentais decorrentes de diferentes aspectos da contabilidade, em suas diversas áreas, tais como, financeira, gerencial e auditoria (Peecher & Solomon, 2001; Libby, Bloomfield, & Nelson, 2002; Sprinkle, 2003).

No contexto brasileiro, estudos demonstram que o uso de experimentos é um fenômeno recente e ainda incipiente (Frezatti, Aguiar, Wanderley e Malagueño, 2015; Homero Jr., 2016). Uma potencial explicação para a pouca atenção dada ao uso de experimentos por pesquisadores brasileiros em contabilidade é a observação de pouca exposição a treinamentos específicos quanto ao uso desse método, seja por meio da oferta de disciplinas nos programas de pós-graduação da área, seja por meio de treinamentos posteriores. Esse pouco conhecimento e experiência com experimentos pode levar àqueles pesquisadores que decidem pelo uso desse método a eventualmente desenhar experimentos que não lidem adequadamente com as potenciais ameaças de validação.

Este artigo discute aspectos centrais que contribuem para que experimentos sejam desenhados e aplicados apropriadamente, incluindo: adequação à questão de pesquisa, operacionalização das variáveis independentes e dependentes, seleção e oferecimento de incentivos aos participantes, e tecnologia para aplicação do estudo. Entende-se que a devida atenção a esses aspectos é essencial para minimizar potenciais problemas de validade, não excluindo, no entanto, atenção adicional para ameaças à validade que sejam específicas a um dado estudo experimental.

Para ilustrar a maneira pela qual esses aspectos são tratados na pesquisa em contabilidade, são usados exemplos de desenhos experimentais publicados em periódicos internacionais da área. A seleção dos artigos usados como exemplos é arbitrária e reflete principalmente meu interesse em aspectos comportamentais associados ao desenho e ao uso de sistemas de incentivos nas organizações, procurando-se incluir estudos de autores com experiência no uso de experimentos.

O artigo é dividido em cinco partes, além desta introdução. Inicialmente, pesquisa experimental é definida e suas principais vantagens e desvantagens são apresentadas. Na sequência, os três principais tipos de experimentos são discutidos. As duas partes seguintes discutem aspectos centrais no desenho e na execução de experimentos de laboratório. Por fim, são elaboradas as considerações finais.

## 2. PESQUISA EXPERIMENTAL – DEFINIÇÃO, VANTAGENS E DESVANTAGENS

Pesquisa experimental é um método de pesquisa quantitativa adequado para verificar relação de dependência entre causas, ou variáveis independentes, e efeitos, ou variáveis dependentes (Shadish, Cook, & Campbell, 2002; Babbie, 2010). Um aspecto particular de experimentos é envolver intervenção

do pesquisador para se verificar mudanças comportamentais decorrentes dessa intervenção (Gall, Gall, & Borg, 2007). Um exemplo de intervenção em um contexto contábil é ao observar a reação dos participantes da pesquisa à presença ou ausência de uma comunicação de valores organizacionais (Kachelmeier, Thornock, & Williamson, 2016) ou à inclusão de diferentes tipos de controle (Christ, Emett, Tayler, & Wood, 2016).

Um apropriado experimento tem três principais propriedades (Leary, 2012): i) pelo menos uma variável independente é manipulada e é verificado os efeitos sobre as respostas dos participantes; ii) participantes são atribuídos às condições experimentais de modo que se garanta que diferentes grupos experimentais sejam equivalentes no maior número possível de aspectos, tais como idade e gênero; e iii) variáveis estranhas e que podem afetar as respostas dos participantes devem ser controladas. As propriedades ii) e iii) visam garantir que o pesquisador poderá atribuir mudanças comportamentais dos participantes, tal como sugerido na primeira propriedade, às manipulações da variável independente. Essa possibilidade de inferência causal está diretamente relacionada à validade interna do estudo.

Experimentos cumprem diferentes propósitos, tais como influenciar decisão de gestores e investidores, identificar regularidades em áreas para as quais não há teoria estabelecida, encontrar regularidades em comportamentos observados e verificar teorias que possam explicar tais regularidades (Friedman & Sunder, 1994). Na pesquisa experimental em contabilidade gerencial, por exemplo, predomina o propósito de influenciar decisão de organizações e gestores quanto ao uso de diferentes estruturas de controle, tais como formais versus informais (Kachelmeier et al., 2016) ou controles de recompensa versus de feedback (Christ et al., 2016).

Uma das vantagens de experimentos é empregar rigor lógico que controle fontes de invalidez, o qual pode ser mais difícil de alcançar por meio de outros métodos (Babbie, 2010). Experimentos podem isolar e controlar efeitos de influências estranhas que limitam ou enviesam as observações, de tal modo que se possa atribuir mudanças na variável dependente ao estímulo experimental (Shadish et al., 2002; Babbie, 2010). Outra vantagem é o de poderem ser replicados com a utilização de diferentes grupos de sujeitos, diferentes contextos e utilizando formas alternativas de mensuração das variáveis dependentes e independentes, o que contribui para fortalecer a validade de construto e a validade externa dos resultados da pesquisa (Shadish et al., 2002; Babbie, 2010).

Dentre as desvantagens no uso de experimentos, destaca-se a artificialidade, já que tarefas experimentais criadas em laboratório podem não necessariamente ocorrer em ambiente social (Babbie, 2010). Além disso, experimentos estão sujeitos a vieses de quem conduz a pesquisa, no sentido de ser impossível o teste de teorias que não reflitam preferências do pesquisador, de maneira que inferências de causa e generalizações podem ser equivocadas (Shadish et al., 2002). Problema adicional com experimentos é que os resultados obtidos, pelo menos em parte, são relativos às suposições e contextos teóricos e podem mudar quando diferentes suposições e contextos são utilizados (Shadish et al., 2002).

### 3. TIPOS DE EXPERIMENTO

É possível distinguir três tipos de experimentos que se complementam em termos de obtenção de um entendimento mais amplo sobre um objeto de pesquisa (Harrison, 2005; Carpenter, Harrison, & List, 2005): natural, campo e laboratório. Cada tipo de experimento tem vantagens e desvantagens relativas, as quais devem ser consideradas na escolha do tipo de experimento (Quadro 1).

Quadro 1: Tipos de experimento: vantagens e desvantagens

Tipos	Vantagens	Desvantagens
Experimento Natural	<ul style="list-style-type: none"> <li>• Reflete escolhas de indivíduos em seu ambiente de decisão;</li> <li>• Maior validade externa em relação aos demais tipos de experimentos.</li> </ul>	<ul style="list-style-type: none"> <li>• Não permite manipulação de variáveis independentes;</li> <li>• Mais exposto ao efeito de variáveis estranhas;</li> <li>• Menor validade interna em relação aos demais tipos de experimento.</li> </ul>
Experimento de Campo	<ul style="list-style-type: none"> <li>• Permite manipulação de variáveis independentes;</li> <li>• Examina tomada de decisão em um ambiente natural;</li> <li>• Maior validade externa em relação a experimentos de laboratório.</li> </ul>	<ul style="list-style-type: none"> <li>• Exposto ao efeito de variáveis estranhas;</li> <li>• Menor validade interna em relação a experimentos de laboratório.</li> </ul>
Experimento de Laboratório	<ul style="list-style-type: none"> <li>• Permite manipulação de variáveis independentes;</li> <li>• Maior controle de variáveis estranhas;</li> <li>• Permite replicação em outros contextos;</li> <li>• Maior validade interna em relação aos demais tipos de experimento.</li> </ul>	<ul style="list-style-type: none"> <li>• Contextos artificiais;</li> <li>• Menor validade externa em relação aos demais tipos de experimento.</li> </ul>

Nota. Elaborado com base nos estudos de Carpenter et al. (2005), Harrison (2005) e Ortmann (2005)

Partindo de experimentos naturais, Shadish et al. (2002) não consideram esse como sendo um tipo de experimento, já que variáveis independentes não podem ser manipuladas. Em um experimento natural, aproveita-se a ocorrência de mudanças exógenas, sobre as quais não se tem controle, e examinam-se os seus efeitos em um desenho experimental (Carpenter et al., 2005; French Jr., 1953). De acordo com Harrison (2005), a principal vantagem de experimentos naturais é refletir escolhas de indivíduos em seu ambiente de decisão, ao invés de considerar situações artificiais normalmente criadas em ambiente de laboratório. Além disso, as implicações dessas escolhas costumam ser significativas e de interesse para certos grupos (Harrison, 2005). Desse modo, os resultados de estudos experimentais têm maior potencial de validade externa. Por sua vez, a sua principal desvantagem é a impossibilidade de se escolher os níveis das variáveis independentes e de se decidir onde e quando será imposto aos indivíduos o efeito dessas variáveis (Harrison, 2005; Carpenter et al., 2005). Além disso, experimentos naturais têm um maior desafio de controle experimental, uma vez que é mais difícil controlar o efeito de variáveis estranhas ao desenho do estudo. Esse aspecto compromete a validade interna do estudo, em termos de estabelecer inferências de causa e efeito.

Exemplo de experimento natural na literatura contábil é o estudo de Gormley, Matsa e Milbourn (2013) o qual examina a relação bidirecional entre remuneração e risco corporativo. No contexto examinado, uma empresa sofre aumento exógeno e não esperado de risco material, permitindo o desenho de experimento natural para investigar como ela ajusta a remuneração gerencial em resposta às mudanças de exposição ao risco e, ainda, como esse ajuste na remuneração impacta a propensão gerencial de investir em projetos mais arriscados. Conforme os autores, a mudança exógena de risco permite um teste da teoria econômica que não seria possível por meio de outro método de pesquisa (Gormley et al, 2013).

Em experimentos de campo, por sua vez, é possível manipular as variáveis independentes (French Jr., 1953), o que vem requerer mais atenção no desenho da pesquisa, em particular, na definição de estratégias de acesso à população alvo (Carpenter et al., 2005). Em relação a experimentos de laboratório, os de campo têm a vantagem de examinar o ambiente natural de tomada de decisões econômicas (Harrison, 2005). Experimentos de campo oferecem ainda a oportunidade de se examinar o efeito de alterar ou implementar novas práticas (p. ex., novo esquema de remuneração) em pequena escala ou piloto, antes de uma total implementação (Carpenter et al., 2005). Enquanto esse tipo de experimento ganha em validade externa ao ser realizado em ambiente natural, resultando em maior nível de generalização, ele também fica mais exposto às questões de validade interna em termos de o quanto é possível atribuir mudanças nas variáveis dependentes às manipulações nas variáveis independentes; em específico, há aumento nas demandas sobre controle experimental no sentido de reduzir o potencial efeito de variáveis estranhas sobre variáveis dependentes (Ortmann, 2005).

Exemplo de experimento de campo em contabilidade é o estudo de Preslee, Vance e Webb (2013) o qual investiga o efeito relativo de recompensas tangíveis e em dinheiro sobre comprometimento com metas e desempenho. Verificam-se duas unidades de uma empresa cujos empregados recebem recompensa em dinheiro pelo alcance de metas e três unidades em que recebem pontos para compra, com valor equivalente à recompensa em dinheiro. Em outro exemplo, Lourenço (2016) examina efeitos complementares e substitutos de tipos de incentivos—monetários, feedback e reconhecimento—sobre desempenho. Em seu estudo, representantes de vendas de uma empresa são aleatoriamente alocados aos grupos experimentais, referentes aos diferentes tipos de incentivos, sem que tenham conhecimento de que diferentes tipos de incentivo estão sendo implantados simultaneamente na empresa.

O tipo de experimento mais comum na literatura contábil, assim como em outras áreas, é o de laboratório. Esse tipo de experimento procura criar situações nas condições exatas pretendidas e em que algumas variáveis são manipuladas e outras são controladas (Festinger, 1953). São então observados e mensurados os efeitos da manipulação das variáveis independentes sobre as variáveis dependentes em um contexto em que se minimiza o efeito de fatores estranhos (Festinger, 1953). Assim, a principal vantagem desse tipo de experimento é o controle que ele permite na condução do estudo, o qual é não contextual e replicável (Harrison, 2005). As principais desvantagens desse tipo de experimento são o fato de normalmente envolverem participantes provenientes de amostras definidas por conveniência, além da natureza abstrata dos ambientes criados em laboratório, o que pode resultar em respostas comportamentais artificiais e não naturais (Ortmann, 2005; Harrison, 2005). Desse modo, em comparação aos outros tipos, experimentos de laboratório ganham em validade interna, enquanto perdem em validade externa.

São muitos os exemplos na literatura contábil de estudos experimentais de laboratório. Alguns exemplos são explorados na sequência deste artigo, a partir da discussão de aspectos centrais no desenho e na execução de experimentos.

## 4. DESENHO DE EXPERIMENTOS DE LABORATÓRIO

### 4.1 Problema de Pesquisa

Experimentos não são adequados para todo e qualquer tipo de estudo; eles são particularmente apropriados para projetos de pesquisa envolvendo conceitos e proposições relativamente limitados e bem definidos (Babbie, 2010). O ponto de partida para o desenho de um experimento é a definição de um problema de pesquisa fundamentado na teoria e a sua adequação a esse problema de pesquisa (French Jr., 1953). O problema de pesquisa deve ser enunciado em termos experimentais, o que significa a necessidade de alto nível de especificidade e clareza em seu enunciado e na definição das variáveis envolvidas (Festinger, 1953).

Por exemplo, Christ et al. (2016) conduzem um experimento de laboratório com o objetivo de investigar, em um ambiente de múltiplas tarefas, se o uso conjunto de controles baseados em recompensa e baseados em feedback pode melhorar desempenho em comparação ao uso isolado de controles baseados em recompensa. Nesse estudo, as variáveis independentes são tipo de controle (recompensa e feedback) e dimensão da tarefa (acurácia e velocidade). Controle baseado em recompensa é definido como um processo que influencia comportamento por meio da remuneração, mas que não fornece feedback imediato de desempenho, tal como um bônus baseado em desempenho. Por sua vez, controle de feedback é definido como um processo que influencia comportamento por meio do feedback de desempenho, mas que não está explicitamente vinculado à remuneração, tal como feedback de

dimensões de desempenho não vinculadas ao pagamento de bônus. Em Christ et al. (2016), o uso do experimento se adequa ao problema de pesquisa na medida em que permite a separação do efeito dos dois tipos de controle, o que seria mais difícil em um ambiente natural.

Uma vez definidos o problema de pesquisa e as variáveis de interesse, torna-se necessária a operacionalização dessas variáveis, de modo alinhado ao problema de pesquisa (Festinger, 1953). É fundamental para a condução bem-sucedida de um experimento que a operacionalização das variáveis de interesse esteja relacionada com a definição conceitual e com a forma pela qual essas variáveis são enunciadas no problema de pesquisa (Festinger, 1953). A vinculação de definições conceituais e operacionais das variáveis independentes é essencial para garantir a validade de constructo de um experimento, permitindo a generalização das relações obtidas entre as definições operacionais das variáveis de interesse para os relacionamentos entre as definições conceituais dessas mesmas variáveis (Maxwell e Delaney, 1990).

Na literatura contábil, é comum o uso do modelo Libby Boxes para discutir a vinculação entre definições conceituais e operacionais de variáveis independentes. Esse modelo é visto como uma descrição útil do processo de teste de hipóteses (Libby, Bloomfield e Nelson, 2002). Sem entrar no detalhe do modelo Libby Boxes, que poderá ser encontrado em Libby et al. (2002), ressalta-se o exemplo de Christ et al. (2016), em que controles baseados em recompensa e baseados em feedback são operacionalizados por sua ausência ou presença, através de informação obtida nas instruções da tarefa experimental. Por exemplo, no caso do grupo experimental em que o controle é baseado em recompensa, a manipulação ocorre por meio da seguinte frase: “Você será compensado ... com base na rapidez e precisão com que você digitar as palavras na tela. Portanto, quanto mais rápido você digitar e quanto mais preciso você digitar, mais dinheiro você vai ganhar!” (Christ et al., 2016, p. 32). Dado o problema de pesquisa e as definições conceituais do estudo de Christ et al. (2016), essa operacionalização permite verificar se a presença dos dois tipos de controle resulta em maior ou menor desempenho em comparação a quando um tipo de controle está presente (recompensa) e o outro ausente (feedback), garantindo assim a vinculação entre definição operacional, conceitual e problema de pesquisa.

Por fim, como forma de identificação prévia de possíveis inadequações na manipulação de variáveis independentes e vinculação entre definições conceituais e operacionais das variáveis e entre essas definições e o problema de pesquisa, é recomendável a realização de sessões experimentais de pré-teste (Festinger, 1953). Essa sugestão é importante quando se admite que a identificação de inadequações somente depois de realizadas as sessões experimentais são mais difíceis de ajustar, ou impossíveis, podendo tornar o experimento inválido. É também comum se incluir questões para validar as manipulações, denominadas de *manipulation checks*, em que se procura determinar se os participantes compreenderam e interpretaram corretamente as variáveis independentes (Libby et al. 2002). Em outras palavras, não se recomenda que se assumam a priori que as manipulações das variáveis independentes foram bem-sucedidas, no sentido de capturarem as definições conceituais de interesse, sendo importante a checagem do nível de sucesso dessa manipulação (Festinger, 1953). Essas questões costumam aparecer depois das instruções para a tarefa experimental ou como parte de questões pós-experimentais.

Essa prática é comum nos estudos em contabilidade. Christ et al. (2016), por exemplo, incluem *comprehension questions* depois das instruções para determinar se os participantes entenderam as manipulações. Em outro exemplo, Koonce, Williamson e Winchel (2010) incluem *manipulation checks* nas questões pós-experimentais para verificar a eficácia das manipulações de seu estudo.

1. “You will be compensated ... based on how quickly and accurately you type the words on the screen. Therefore, the faster you type, and the more accurate your output, the more money you will earn”.

## 4.2 Variáveis Independentes

As variáveis independentes cumprem papel fundamental em experimentos de laboratório. Com o propósito de responder ao problema de pesquisa, o passo inicial é especificar os fatores cujos efeitos se pretende estudar (Myers, 1979). Assim, um primeiro passo pode ser a definição da quantidade de variáveis independentes que serão incluídas em um estudo. A princípio, não há restrições de quantidade, embora seja importante considerar que quanto maior o número de variáveis independentes, maior a complexidade do estudo e maior o número de participantes necessários. Em geral, o mais comum é a inclusão de duas variáveis independentes, inclusive como forma de se tirar proveito da vantagem adicional de experimentos de laboratório de permitir a análise de efeitos de interação/moderação entre variáveis independentes.

No exemplo de Christ et al. (2016), examina-se a interação de duas variáveis independentes. Kachelmeier et al. (2016) incluem duas variáveis independentes para verificar como a presença de uma declaração de valores modera o efeito do esquema de incentivo sobre a produtividade no desempenho da tarefa experimental. Libby, Salterio e Webb (2004), por sua vez, incluem três fatores: *accountability* na avaliação de desempenho, relatório de *assurance* sobre confiabilidade de resultados e avaliação de desempenho de divisões. O predomínio de dois fatores não significa que existam restrições a experimentos com apenas uma variável independente (ver p. ex., Hannan, Rankin & Towry, 2010; Choi, Hecht, Tafkov & Towry, 2016). Também se destaque que existem estudos envolvendo múltiplos períodos ou *rounds* em que o número de períodos ou *rounds* é indicado como variável independente para fins de desenho experimental (ver p. ex., Hannan et al., 2010; Guo, Libby & Liu, 2017).

Uma vez definido o número de variáveis independentes, torna-se necessária a definição operacional dessas variáveis, o que pode ocorrer por manipulação e/ou mensuração (Quadro 2). Em um experimento, pelo menos um fator tem que ser manipulado, isto é, os níveis de operacionalização da variável independente são diferentes para diferentes grupos experimentais, para que se possa verificar os efeitos de mudanças nesses níveis sobre o comportamento dos participantes (Leary, 2012). Quando não há variável manipulada, o estudo é caracterizado como quase-experimental (Leary, 2012). Mass, Van Rinsum e Towry (2012), por exemplo, conduzem um quase-experimento em que o fator, desempenho agregado, é mensurado ao invés de manipulado.

A mensuração de variáveis independentes em um experimento é tipicamente feita por meio de medidas de percepção com o objetivo de capturar pensamentos, sentimentos ou comportamentos (Leary, 2012). Embora não sendo uma regra, em geral, as variáveis mensuradas em um experimento são usadas como moderadoras ou mediadoras em razão da dificuldade em se manipular algumas dessas variáveis (Libby et al., 2002). A mensuração de variáveis independentes normalmente aparece em questões pós-experimentais, depois da mensuração das variáveis dependentes. Um experimento, portanto, pode utilizar apenas variáveis manipuladas ou variáveis manipuladas e mensuradas, quando mais de uma variável independente é incluída no desenho experimental, ou quando variáveis moderadoras e mediadoras são de interesse. Seja qual for a escolha, recomenda-se que a operacionalização considere estudos anteriores similares ou alinhamento teórico (Myers, 1979).

Como exemplo de variável mensurada, o estudo de Christ, Sedatole e Towry (2012) examina o efeito *framing* do contrato de incentivo sobre o nível de esforço em um contexto de contratos incompletos. Christ et al. (2012) incluem dois fatores: representação do contrato e implementação do contrato. O

primeiro é manipulado em dois tipos, contrato de bônus e de penalidade. O segundo fator é mensurado em dois tipos, indicando a decisão do principal de implementar ou não o contrato de incentivo escolhido (Christ et al., 2012). Em outro exemplo, Kachelmeier et al. (2016) mensuram o número de respostas incorretas à tarefa experimental e utilizam essa variável como mediadora do efeito do tipo de incentivo sobre produtividade.

Quadro 2: Desenho de experimentos de laboratório: variáveis independentes

Aspectos	Dimensões	Detalhamento
Definição operacional	Manipulação	<ul style="list-style-type: none"> <li>• Necessário pelo menos uma manipulação para caracterizar um experimento;</li> <li>• Pode ser em termos quantitativos (níveis diferem em montante) ou qualitativos (níveis diferem em tipo).</li> </ul>
	Mensuração	<ul style="list-style-type: none"> <li>• Normalmente por meio de medidas de percepção;</li> <li>• É geralmente incluída como variável moderadora ou mediadora;</li> <li>• Aparece nas questões pós-experimentais.</li> </ul>
Desenho experimental	<i>Between-participants</i>	<ul style="list-style-type: none"> <li>• Quando o propósito é entender diferenças em comportamento entre diferentes grupos de participantes;</li> <li>• Cada grupo experimental é exposto a apenas uma condição experimental (um nível da variável independente);</li> <li>• Requer maior número de participantes;</li> <li>• Recomenda-se aleatorização dos participantes às condições experimentais.</li> </ul>
	<i>Within-participants</i>	<ul style="list-style-type: none"> <li>• Quando o propósito é entender diferenças em comportamento através das condições experimentais em relação ao mesmo grupo de participantes;</li> <li>• Cada participante do estudo é exposto a todas as condições experimentais;</li> <li>• Maior poder para identificação do efeito dos fatores (variáveis independentes);</li> <li>• Requer menor número de participantes;</li> <li>• Cada participante serve como seu próprio controle experimental.</li> </ul>
	Misto	<ul style="list-style-type: none"> <li>• Inclui pelo menos uma variável manipulada <i>between-participants</i> e pelo menos uma manipulada <i>within-participants</i>;</li> <li>• Recomenda-se aleatorização se a manipulação <i>between-e within-participants</i> não for em relação à mesma variável.</li> </ul>
	Hierárquico	<ul style="list-style-type: none"> <li>• Uma variável é manipulada aninhada aos níveis de outra variável;</li> <li>• Não adequado para exame de efeitos interativos entre duas ou mais variáveis;</li> <li>• Recomenda-se aleatorização.</li> </ul>

A variável independente pode ser manipulada em termos quantitativos ou qualitativos (Myers, 1979; Leary, 2012). A manipulação quantitativa é aquela cujos níveis diferem em montante, de maneira a se verificar o efeito de mudanças na quantidade manipulada sobre mudanças no comportamento. Guo et al. (2017) oferecem exemplo de operacionalização quantitativa da manipulação. Nesse estudo, a variável independente é dispersão vertical da remuneração, ou seja, diferença na remuneração entre diferentes níveis hierárquicos dentro de uma organização. Os autores consideram três montantes para essa variável que diferem em termos do quociente de salário fixo entre superior e subordinado: \$25:\$25, \$25:\$10 e \$10:\$10.

Por sua vez, manipulação qualitativa é aquela cujos níveis diferem em tipo, de modo a se verificar o efeito de tipos específicos da variável independente sobre escolhas dos indivíduos (Myers, 1979). A operacionalização da variável ‘tipo de controle’ no estudo de Christ et al. (2016) é um exemplo de manipulação qualitativa. Em seu estudo, os dois tipos de controle são mensurados em termos de presença ou ausência e é verificado o seu efeito sobre desempenho.

Um aspecto adicional relativo às variáveis independentes se refere à definição do tipo de desenho. Destacam-se quatro principais tipos de desenho experimental: *between*, *within*, misto e hierárquico (Quadro 2). Se o propósito do estudo é entender diferenças em comportamento entre diferentes grupos de participantes, esses diferentes grupos são atribuídos a apenas uma condição experimental, caracterizando um desenho *between-participants* (Leary, 2012). Na literatura contábil, predominam desenhos *between-participants* (ver p. ex., Kachelmeier et al., 2016; Christ et al., 2016; Christ et al., 2012), em que, de modo geral, o objetivo é comparar o comportamento entre diferentes indivíduos expostos a diferentes níveis das variáveis independentes.

Quando, no entanto, o propósito é entender diferenças em comportamento através das condições experimentais em relação ao mesmo grupo de participantes, esse grupo é exposto a todas as condições,

caracterizando um desenho *within-participants*, também conhecido como desenho de mensurações repetidas (*repeated measures design*) (Leary, 2012). Em um experimento de laboratório de mercado, Bloomfield e Wilks (2000) oferecem um exemplo em que manipulam três variáveis independentes *within-participants*—qualidade de evidenciação de uma ação, qualidade de evidenciação de outras ações e choques de demanda.

Ao se comparar esses dois tipos de experimento, estudos *within-participants* são considerados superiores a desenhos *between-participants* por serem mais poderosos, ou seja, por serem mais capazes de identificar efeitos dos fatores, e ainda por requererem menos participantes, uma vez que cada participante é exposto a todas as condições experimentais (Leary, 2012). Além disso, desenhos *within-participants* prescindem da necessidade de aleatoriedade na atribuição de participantes às condições experimentais, uma vez que cada participante serve como seu próprio controle ao participar de todas as condições (Leary, 2012). Entretanto, experimentos *within-participants* também têm desvantagens, tais como, melhoria no desempenho dos sujeitos em razão da prática seguida com a tarefa experimental, cansaço durante o experimento, percepção da hipótese de pesquisa em investigação, e permanência do efeito de um nível da variável independente quando os participantes estão sendo expostos ao segundo nível dessa variável (*carryover effects*) (Leary, 2012). Diante de vantagens e desvantagens relativas, a escolha entre esses dois tipos de desenho experimental deveria ser direcionada pelo problema de pesquisa e pelos recursos à disposição (p. ex., acesso a um maior número de participantes).

Desenhos mistos incluem pelo menos uma variável manipulada *between-participants* e pelo menos uma manipulada *within-participants*. Libby et al. (2004) oferecem um exemplo em que manipulam duas variáveis *between-participants*—*accountability* na avaliação de desempenho e relatório de assurance sobre confiabilidade dos resultados—e uma variável *within-participants*—avaliação de desempenho de divisões. Cloyd, Pratt e Stock (1996) adotam desenho misto, sendo que as mesmas variáveis são manipuladas *between-* e *within-participants* com o objetivo de aumentar a validade dos resultados. Importante mencionar que, em estudos envolvendo múltiplos períodos ou rounds (ver p. ex., Guo et al., 2017; Hannan et al., 2010), a variável período ou round é normalmente operacionalizada *within-participants*, uma vez que o mesmo participante é exposto a essa variável independente ao longo de toda a sessão experimental.

Por fim, em desenhos hierárquicos, uma variável é manipulada aninhada aos níveis de outra variável, o que impede o exame de efeitos interativos entre as duas variáveis (Myers, 1979; Leary, 2012). Koonce et al. (2010) oferecem exemplo em que examinam o papel de divulgações *ex post* da acurácia de estimativas contábeis no comportamento de investidores não profissionais. Dois fatores são manipulados. O primeiro, se a evidenciação revela estimativas precisas ou não, é manipulado *between-participants*. O segundo fator, tipo de consenso, é também manipulado *between-participants*, aninhado a apenas uma das condições da primeira variável, no caso, quando a evidenciação de estimativas é não precisa. São considerados três níveis da variável aninhada: alto, baixo ou sem informação. Logo, o estudo de Koonce et al. (2010) resulta em quatro condições: preciso, não preciso-alto consenso, não preciso-baixo consenso, e não preciso-sem informação de consenso.

Com exceção de experimentos *within-participants*, os demais tipos requerem a definição de um método de atribuir os sujeitos às condições experimentais criadas pelas variáveis independentes. Para ser caracterizado como experimento, deve se realizar a atribuição por meio de aleatorização (Leary, 2012), em que cada sujeito é alocado a uma das condições experimentais através de processo aleatório, tornando os grupos mais homogêneos em relação ao maior número possível de fatores e contribuindo assim para a validade interna do estudo (Shadish et al., 2002). Quando não houver essa aleatorização,

o estudo é dito como sendo um quase-experimento (Shadish et al., 2002). Vale destacar ainda que a simples aleatorização representa um importante processo para eliminar diversas ameaças de validade interna, em particular, à ameaça de seleção<sup>2</sup> (Shadish et al., 2002).

Aspecto adicional no desenho experimental e que também contribui para sua validade interna é o uso de grupos de controle, além de grupos experimentais. Enquanto grupos experimentais são expostos aos níveis das variáveis independentes, grupos de controle não sofrem essa mesma exposição (Gall et al., 2007), ou são expostos a certo nível das variáveis independentes que serve como base de comparação com as demais condições experimentais (Leary, 2012). A combinação entre grupos de controle e aleatorização dos participantes às condições experimentais aumenta a validade interna da inferência causal, sendo possível dizer que diferenças na mensuração da variável dependente entre os grupos de controle e os experimentais são atribuídas à exposição aos níveis das variáveis independentes, uma vez que os dois grupos são equivalentes (Babbie, 2010). Importante também a inclusão de questões demográficas, como parte de questões pós-experimentais, para se verificar a similaridade entre grupos experimentais e de controle.

Em sua maioria, estudos experimentais em contabilidade utilizam grupos de controle. No exemplo de Christ et al. (2016), em uma das condições experimentais, os sujeitos não são expostos a qualquer dos tipos de controle (recompensas ou feedback), servindo a produtividade desses sujeitos (grupo de controle) como base de comparação (baseline) com a produtividade dos demais grupos experimentais.

Os aspectos discutidos caracterizam o experimento, indicando se o mesmo é, por exemplo,  $2 \times 2$  *between-subjects* ou  $2 \times 2$  *between-subjects*  $\times$   $10$  *within-subjects*. O estudo de Christ et al. (2016) é  $2 \times 3$  *between-subjects*, pois um fator (dimensão da tarefa) é manipulado em dois níveis, acurácia e velocidade, e o outro fator (tipo de controle) é manipulado em três níveis, controle de recompensas, feedback, ou sem controle. Em Hannan et al. (2010), o estudo é  $2$  (*between-subjects*)  $\times$   $8$  (*within-subjects*), já que um fator, amplitude de controle, é manipulado em dois níveis, alto e baixo, e todos os sujeitos são expostos a oito períodos da tarefa experimental. Como exemplo final, o estudo de Libby et al. (2004) é caracterizado como  $2 \times 2 \times 2$ , em que um fator é manipulado *within-subjects* em dois níveis e os dois outros fatores são manipulados *between-subjects*, cada qual também em dois níveis.

## 4.3 Variáveis Dependentes

As variáveis dependentes em um experimento são aquelas mensuradas e em relação às quais se espera verificar diferenças em razão de manipulações nos níveis das variáveis independentes (Leary, 2012). Um estudo pode selecionar uma ou mais variáveis dependentes, conforme o problema de pesquisa investigado. Idealmente, espera-se que tais variáveis possuam determinadas características as quais podem afetar a validade do estudo, incluindo confiabilidade e sensibilidade (Myers, 1979). A operacionalização da variável dependente normalmente envolve observação de comportamento e uso de medidas de percepção (Leary, 2012). Assim como ocorre com variáveis independentes, é importante se considerar as potenciais ameaças à validade de construto na operacionalização de variáveis dependentes.

Na literatura contábil, Christ et al. (2016) oferecem um exemplo em que mensuram duas variáveis dependentes por observação: precisão e velocidade em inserir dados. Christ et al. (2016) definem ainda uma medida agregada de desempenho a partir do desempenho nas duas dimensões da tarefa experimental. O estudo de Kachelmeier et al. (2016) oferecem outro exemplo em que a principal variável dependente, produtividade (*correct output*), é mensurada por meio de observação, embora seja

2. Shadish et al. (2002) apresentam uma lista de ameaças aos diferentes tipos de validade e potenciais remédios para reduzir o efeito de tais ameaças.

também considerada uma segunda variável, número de respostas incorretas, a qual é posteriormente utilizada como mediadora. No estudo de Choi et al. (2016), duas das três variáveis dependentes, distância psicológica e nível de construção, são mensuradas por meio de escalas de percepção.

## 5. EXECUÇÃO DE EXPERIMENTOS DE LABORATÓRIO

### 5.1 Participantes

Um relevante aspecto na execução de experimentos e que está relacionado à sua validade externa é a decisão sobre os participantes, em termos de quem serão (de qual população serão retirados) e em termos do número total necessário, dado o número de grupos experimentais (Festinger, 1953; Myers, 1979). Quanto ao primeiro aspecto, a escolha dos participantes é normalmente direcionada pelo propósito da pesquisa (Myers, 1979). Na prática, predomina escolha de participantes padronizados ou homogêneos uma vez que são recrutados a partir de amostras de estudantes definidas por conveniência (Carpenter et al., 2005). A vantagem na escolha de participantes homogêneos é a de reduzir a possibilidade de mudanças na variável dependente decorrentes de diferenças no perfil dos participantes, o que contribui para a validade interna do estudo (Trochim & Donnelly, 2008). Além disso, a homogeneidade dos participantes reduz o erro de estimação do efeito da variável independente sobre a dependente, contribuindo para um maior poder dos resultados e maior validade de conclusão ou estatística (Shadish et al., 2002).

Por outro lado, em razão do uso de participantes definidos por conveniência, além da natureza abstrata de tarefas experimentais, é que se questiona a validade externa de estudos de laboratório (Maxwell e Delaney, 1990; Ortmann, 2005). Para lidar com essa crítica, experimentos de laboratório têm recrutado participantes não padronizados, mais heterogêneos e que estejam no campo (Ortmann, 2005), tal como investidores profissionais (ver p. ex., Elliot, Hodge, Kennedy e Pronk, 2007). Estudos incluindo participantes não padronizados são denominados por Carpenter et al. (2005) de experimentos de campo artefactuais, tendo por única diferença vis-à-vis estudos convencionais de laboratório o fato de usar participantes não padronizados. O contraponto no uso de participantes mais heterogêneos é o aumento da possibilidade de que diferenças existentes entre os participantes estejam relacionadas com a variável dependente, dificultando a inferência causal e, portanto, comprometendo a validade interna e a validade de conclusão do estudo (Trochim & Donnelly, 2008; Shadish et al., 2002).

Diferenças significativas preexistentes entre os participantes podem permanecer mesmo depois da aleatorização aos grupos experimentais, sendo uma forma de lidar com essa potencial ameaça a de comparar os grupos experimentais no maior número possível de variáveis de perfil e, em se identificando diferenças significativas, incluir formalmente as variáveis que indicam diferenças como covariáveis no modelo de análise (Shadish et al., 2002). Outra alternativa para lidar com essas diferenças é desenhar um estudo experimental que inclua pré-testes das mesmas variáveis de desempenho que se pretende mensurar posteriormente à manipulação das variáveis independentes e incluir esse resultado no processo de análise (Shadish et al., 2002).

Ainda em relação à seleção dos participantes, Myers (1979) reconhece que certos processos básicos não parecem requerer, pelos menos em uma análise inicial, um conjunto específico de participantes. Em um contexto contábil, Libby et al. (2002) destacam não ser necessária a seleção de participantes com nível de sofisticação maior do que aquele demandado pela tarefa experimental. Além disso, Elliot et al. (2007) indicam que participantes não padronizados (investidores profissionais) e participantes

padronizados (estudantes de pós-graduação lato sensu em diferentes estágios do curso) são similares em tarefas experimentais de baixa complexidade integrativa; por sua vez, em tarefas de alta complexidade, participantes padronizados continuam sendo similares aos não padronizados, mas apenas quando estiverem em estágios mais avançados do curso de pós-graduação.

Tipicamente, experimentos de laboratório em contabilidade utilizam participantes padronizados, representados por estudantes, de graduação ou pós-graduação. Por exemplo, Christ et al. (2016) justificam o uso de estudantes de graduação em razão da tarefa experimental (digitação de dados) se adequar às habilidades de um típico estudante de graduação. Koonce et al. (2010), por sua vez, utilizam estudantes de pós-graduação lato sensu, com experiência anterior de investimento no mercado de ações e com conhecimentos razoáveis na área de negócios e economia. Koonce et al. (2010) justificam a escolha por participantes padronizados por considerarem que os mesmos se adequam aos propósitos do estudo e por atuarem como investidores não profissionais. Por fim, Cloyd et al. (1996) oferecem um exemplo de utilização de participantes não padronizados, ao selecionar especialistas tributários de empresas industriais norte-americanas em seu desenho experimental.

O segundo aspecto relevante relativo aos participantes se refere ao número total de participantes. Sugere-se que a decisão do número de participantes se faça com base no poder desejado, ou seja, na probabilidade desejada de concluir corretamente que existem diferenças entre os efeitos provocados pelas variáveis manipuladas sobre as variáveis dependentes (Myers, 1979). A decisão do número de participantes tem, portanto, um efeito direto na validade de conclusão ou estatística do estudo (Shadish et al., 2002; Trochim & Donnelly, 2008). Como regra geral, quanto maior o número de sujeitos, maior o poder de um determinado desenho experimental (Leary, 2012). De modo mais prático, o mais comum em experimentos é a busca por, no mínimo, 20 participantes por condição experimental, buscando-se normalmente tamanhos balanceados por grupo experimental, ou seja, com igual número de participantes (Smith, 2000).

Assim, se um experimento é do tipo  $2 \times 2$  *between-subjects*, resultando em quatro condições experimentais, seriam necessários, pelo menos, 80 sujeitos (20 sujeitos  $\times$  4 condições experimentais). Kachelmeier et al. (2016), por exemplo, recrutam participantes de modo a ter 20 em cada condição experimental de um estudo  $2 \times 2$  *between-subjects*, de maneira a totalizar 80 participantes. Christ et al. (2016), em seu estudo  $2 \times 3$  *between-subjects* (incompleto), recruta 125 sujeitos, com quantidade aproximada de 25 sujeitos por condição experimental. Conforme já mencionado, uma vantagem de experimentos do tipo *within-subject* é a necessidade de um número menor de participantes, uma vez que o mesmo participante é exposto a todas condições experimentais. Desse modo, se um experimento é do tipo  $2 \times 2$  *within-subjects*, esse estudo requer metade do número de participantes que seria necessário caso o estudo fosse *between-subjects*.

## 5.2 Escolha da Tecnologia

Uma decisão a ser tomada na execução de experimentos é a tecnologia por meio da qual os participantes serão expostos às condições experimentais. Existem três principais alternativas: *pencil-and-paper*, computador e internet. A primeira é a mais tradicional em que todo o procedimento experimental é realizado por meio de papel e os participantes anotam suas respostas comportamentais às manipulações do estudo. Essa alternativa tem a vantagem de ser um método de coleta de dados confiável e que requer mínimo conhecimento de tecnologia (Bryant, Hunton e Stone, 2004). Por sua vez, as desvantagens dessa alternativa incluem: limitação de acesso a amostras maiores, impossibilidade ou dificuldade de maior interatividade nas respostas, maior probabilidade de os participantes se comportarem tal como entendem ser esperado (*demand effects*), e maior esforço de tabulação de dados (Reips, 2000).

Em particular, o fato de ser mais provável a ocorrência do *demand effects* é um aspecto que compromete a validade interna do estudo.

Exemplo de experimento em contabilidade realizado por meio de *pencil-and-paper* é o de Choi et al. (2016) que separam os materiais do experimento em dois envelopes: o primeiro contém o caso hipotético para apresentação das condições experimentais e o segundo contém questões demográficas e de checagem. Como mencionam, é recomendado aos participantes abrirem o segundo envelope somente depois de completarem as tarefas previstas no primeiro envelope (Choi et al., 2016).

A segunda alternativa de tecnologia é por meio do uso de computador. Esse é certamente o meio mais comum, através do qual, experimentos de laboratório em contabilidade são atualmente executados, existindo softwares destinados para essa finalidade, tal como o Z-Tree. A principal vantagem dessa tecnologia é possibilitar a obtenção de melhores controles experimentais, quanto à aleatoriedade e controle do tempo, tanto de exposição às manipulações experimentais quanto de realização do experimento (Bryant et al., 2004). Além disso, essa alternativa permite interatividade em tempo real no processo de coleta de dados, facilitando identificação imediata de respostas às manipulações experimentais (Bryant et al., 2004). Como desvantagens dessa alternativa, destaquem-se necessidade de maior conhecimento em tecnologia e maiores custos envolvidos na programação e coleta de dados (Bryant et al., 2004).

Exemplos de experimentos em contabilidade realizados em laboratório com computador são diversos, incluindo os estudos já mencionados de Kachelmeier et al. (2016), Guo et al. (2017) e Christ et al. (2016). Destaque-se que, no estudo de Kachelmeier et al. (2016), utiliza-se *pencil-and-paper* para apresentar a tarefa experimental e computador para registrar as respostas à tarefa experimental. Em Kachelmeier et al. (2016), o uso de computador oferece a vantagem de permitir *feedback* imediato aos participantes quanto a ter sido a resposta correta ou incorreta, o que representa um componente importante do propósito da pesquisa.

É ainda possível a aplicação do experimento via internet em que a tarefa experimental é encaminhada por mensagem eletrônica aos potenciais participantes. A principal vantagem no uso de internet é o alcance de amostras de participantes ainda maiores e mais heterogêneas (Leary, 2012). Além disso, por não existirem pesquisadores presentes na aplicação do experimento, há menor probabilidade de os participantes se comportarem tal como entendem que é esperado (Leary, 2012). Dentre as desvantagens no uso de internet, destacam-se: dificuldade em se confirmar a identidade dos participantes, dificuldade de controle do ambiente em que o experimento está sendo realizado e ainda maior possibilidade de desistência de participantes ao longo da aplicação do experimento (Leary, 2012). Em conjunto, essas desvantagens representam ameaças à validade interna do estudo. Bryant et al. (2004) discutem aspectos adicionais de qualidade dos dados e validação na condução de experimentos via internet.

O estudo de Kelly (2007) é um exemplo de experimento realizado por meio de internet em que os participantes acessam o experimento com usuário e senha que lhes são específicos, o que reduz o problema de identificação dos participantes. Bol e Smith (2011) é outro exemplo em que, com ajuda do programa usado para coleta de dados, atribuem aleatoriamente os sujeitos às condições experimentais, o que reduz problemas de controle experimental.

Por fim, destaquem-se tecnologias alternativas utilizadas em experimentos de laboratório. Uma dessas alternativas, mais comum em estudos de contabilidade com foco em variáveis de mercado, é o uso de experimentos em laboratório de mercado (Bryant et al., 2004). O estudo de Bloomfield e Wilks (2000) oferece exemplo do uso dessa tecnologia. Outra alternativa é também conduzida via internet, diferindo na forma de recrutamento de participantes. Trata-se do *Amazon Mechanical Turk* que, ainda que recente, tem se tornado popular em experimentos em ciências sociais (Paolacci, Chandler e Ipeirotis, 2010). Chen, Pesch e Wang (2016) e Christ e Vance (2015) são exemplos recentes em contabilidade usando essa tecnologia.

### 5.3 Uso de Incentivos Monetários

Uma das principais críticas a experimentos de laboratório é a artificialidade das condições experimentais. Em particular, uma crítica reside no questionamento de se determinado comportamento obtido em laboratório aconteceria em ambiente natural em que perdas e ganhos ‘reais’ estivessem em jogo (Carpenter et al., 2005). Além disso, o uso de incentivos monetários em experimentos de laboratório tem sido indicado em razão de seu efeito motivador sobre os participantes, quanto a engajamento com as tarefas experimentais (Friedman e Sunder, 1994). Logo, uma forma adicional de reduzir a criticada artificialidade de experimentos de laboratório seria por meio do oferecimento de incentivos monetários aos participantes.

Entretanto, importante mencionar discussão existente no âmbito da literatura sobre motivação quanto ao efeito negativo sobre motivação intrínseca ao se oferecer incentivos extrínsecos (Deci, 1971). Gneezy e Rustichini (2000) testam o efeito de incentivos monetários em um experimento, indicando que aqueles que não recebem incentivos monetários apresentam desempenho superior aos que recebem baixo incentivo monetário. Camerer e Hogarth (1999) destacam ainda que experimentos divergem quanto ao efeito de incentivos monetários no desempenho: de um lado, presença e montante do incentivo afetam desempenho médio em muitas tarefas, em particular, quando envolvem julgamento; de outro lado, em relação a tarefas, tais como previsão e solução de problemas, incentivos podem reduzir desempenho.

Na literatura experimental contábil, enfatiza-se que a decisão de oferecer ou não incentivos monetários deveria ser direcionada pelos propósitos do estudo (Libby et al., 2002). Bonner e Sprinkle (2002) indicam ainda que estudos contábeis feitos em laboratório divergem quanto ao efeito de incentivos monetários no desempenho; sugerindo a necessidade de estudos que procurem entender fatores que moderam a eficácia no uso de incentivos monetários. Dentre os fatores moderadores, Bonner, Hastie, Sprinkle e Young (2000) sugerem que tipo de tarefa experimental e tipo de esquema de incentivo afetam a eficácia dos incentivos monetários, de modo que incentivos monetários vinculados a tarefas de maior complexidade cognitiva e que requerem maior nível de habilidade não aumentam desempenho dos sujeitos. Além disso, estudos em que esquemas de incentivo são um dos fatores, a sua inclusão é componente essencial no sentido de ser verificado o efeito do incentivo monetário sobre o comportamento dos participantes (Bonner et al., 2000).

Em Kachelmeier et al. (2016), por exemplo, um de seus fatores é o tipo de incentivo, o qual é considerado em dois níveis ou tipos: *piece-rate* e fixo. Participantes expostos ao tipo *piece-rate* são remunerados com base no número de respostas corretas registradas no sistema, de modo que quanto maior o número de respostas corretas, maior a remuneração que pode ser recebida. Participantes expostos ao esquema fixo, por sua vez, recebem valor fixo por participarem. Em média, o montante pago aos participantes nos diferentes tipos de incentivo é aproximadamente igual (Kachelmeier et al., 2016), o que é importante para que variações nas variáveis dependentes não sejam atribuídas a variações no montante de remuneração, mas apenas ao tipo de remuneração.

Uma vez que a decisão em determinado desenho experimental seja a de usar incentivos, questão adicional que emerge é o valor do incentivo para que de fato ele cumpra com o seu papel. Enquanto alto montante de incentivo monetário resulta em maior desempenho dos participantes em relação a baixo montante, montantes ainda mais altos não resultam em incrementos no desempenho (Gneezy e Rustichini, 2000). De fato, é possível que o pagamento de valores bastante altos resultem em redução no desempenho (Ariely, Gneezy, Loewenstein, e Mazar, 2009). Em contabilidade, Libby et al. (2002) indicam que a maioria dos experimentos verificando teorias econômicas pagam aos participantes valor medio de US\$8 a US\$20 por hora, com pagamentos feitos no intervalo entre US\$5 a US\$100 por hora,

em alguns casos. Em Kachelmeier et al. (2016), por exemplo, os participantes recebem uma média de US\$23 de remuneração.

Em síntese, a decisão do valor a ser usado como remuneração depende do contexto em que a pesquisa está sendo realizada. Por exemplo, um valor de R\$20 pode ser significativo para a participação de estudantes de graduação, enquanto que o mesmo não pode ser dito se os participantes forem executivos. Destaque-se ainda a existência de alternativas para remunerar participantes em um experimento. Guo et al. (2017), por exemplo, ao invés de remunerar com base em desempenho acumulado, escolhem aleatoriamente o desempenho obtido em uma das rodadas do experimento para realizar o pagamento aos participantes. Em outro exemplo, Bol e Smith (2011) oferecem um incentivo monetário definido de forma aleatória, sendo que agora não havendo qualquer relação com o desempenho dos participantes no estudo. Outros estudos utilizam uma moeda fictícia que depois é convertida em uma moeda real com base em alguma taxa, permitindo aumentar a saliência dos valores sem onerar excessivamente a pesquisa. Hecht, Tafkov e Towry (2012), por exemplo, utilizam ‘lira’ como a moeda do estudo, sendo convertida para dólares norte-americanos ao final do experimento a uma taxa de 240 ‘liras’ por dólar.

Uma alternativa a incentivos monetários é o uso de vale-presentes como um incentivo tangível com valor monetário (ver p. ex., Choi et al., 2006). A concessão de créditos como forma de incentivo também é uma alternativa, em especial, quando os participantes são estudantes de graduação (ver p. ex., Pinsker, 2011; Tan, Wang e Zhou, 2015). Por fim, importante separar pagamento de incentivos por desempenho obtido na tarefa experimental e pagamento por simples comparecimento para participação no estudo, ou taxa de participação (ver p. ex., Kuang e Moser, 2011; Hannan, Towry e Zhang (2013). Enquanto o primeiro está associado diretamente ao envolvimento do participante no estudo, o segundo pode ser realizado mesmo que o participante não finalize o estudo, sendo normalmente denominado de *show-up fee*.

Quadro 3: Execução de experimentos de laboratório

Aspectos	Dimensões	Detalhamento
Participantes	Definição do perfil	<ul style="list-style-type: none"> <li>• Homogêneos – contribui para validade interna a custo de menor escopo de generalização;</li> <li>• Heterogêneos – contribui para validade externa, criando, no entanto, potencial para diferenças na variável dependente em razão de diferenças preexistentes no perfil dos participantes;</li> <li>• Nível de sofisticação dos participantes depende das demandas exigidas pela tarefa experimental (Libby et al., 2002).</li> </ul>
	Definição da quantidade	<ul style="list-style-type: none"> <li>• Um maior número de participantes por condição experimental contribui para aumentar o poder do estudo e sua validade de conclusão ou estatística.</li> </ul>
Tecnologia	<i>Pencil-and-paper</i>	<ul style="list-style-type: none"> <li>• Participantes anotam em papel suas respostas comportamentais;</li> <li>• Fácil entendimento e aplicação;</li> <li>• Pode não ser adequado se intenção for obter maiores amostras ou se houver interesse em maior interatividade nas respostas.</li> </ul>
	Computador	<ul style="list-style-type: none"> <li>• Uso de softwares ou programação para realização do experimento;</li> <li>• Implementação mais eficiente de controles experimentais (p. ex., aleatorização);</li> <li>• Permite interatividade em tempo real na coleta de dados;</li> <li>• Pode ser mais custoso e demandar mais tempo.</li> </ul>
	Internet	<ul style="list-style-type: none"> <li>• Tarefa experimental é encaminhada via mensagem eletrônica a potenciais participantes;</li> <li>• Pode alcançar um número elevado e diversificado de participantes;</li> <li>• Maior dificuldade de realizar controles experimentais.</li> </ul>
Incentivos Monetários	Definição da utilização	<ul style="list-style-type: none"> <li>• Existem controvérsias teóricas e empíricas quanto à efetividade;</li> <li>• Deve ser utilizado se for um dos fatores.</li> </ul>
	Definição do valor	<ul style="list-style-type: none"> <li>• Valor pode depender de diferentes aspectos do contexto da pesquisa, tal como renda média dos participantes;</li> <li>• Alternativas; escolha aleatória da rodada experimental para definição da remuneração, sorteio, uso de moeda fictícia, uso de vale-presentes e concessão de créditos</li> </ul>

Em síntese, diferentes incentivos são usados em experimentos de laboratório com o objetivo de aumentar a saliência da tarefa experimental e engajar os sujeitos. Criatividade na definição de qual incentivo utilizar é essencial para haver adequação às restrições orçamentárias, sem prejuízo de adequação aos propósitos do estudo. O Quadro 3 sumariza os principais aspectos discutidos em relação à execução de experimentos de laboratório.

## 6. COMENTÁRIOS FINAIS

O propósito deste estudo foi destacar aspectos centrais no desenho e na execução de experimentos de laboratório os quais contribuem para o uso apropriado desse método de pesquisa em termos de obtenção de resultados válidos. Os aspectos explorados neste estudo, embora importantes, estão longe de esgotar a variedade existente de aspectos no desenho de estudos experimentais. Por exemplo, pouca ou nenhuma atenção foi dada a aspectos, tais como inclusão de grupos de pré-teste, inclusão de múltiplos pós-testes para as variáveis dependentes, inclusão de variáveis dependentes não equivalentes, dentre outros aspectos de desenho de experimentos destacados em Shadish et al. (2002), os quais contribuem para mitigar determinadas ameaças de validação, em particular, validação interna. De modo semelhante, este estudo não entrou em detalhes de questões específicas relativas ao desenho de quase-experimentos e modelos relevantes de desenhos quase-experimentais, tais como o desenho de regressão descontínua (Shadish et al., 2002; Trochim & Donnelly, 2008). Ao se negligenciar esses aspectos adicionais no desenho de experimentos, a intenção não foi de indicar que tais aspectos não sejam importantes, mas apenas que os aspectos centrais discutidos neste estudo potencialmente atingem um maior número de desenhos experimentais, enquanto aqueles podem ser mais específicos a um ou outro desenho experimental.

Embora sendo relevante em qualquer estudo experimental, este estudo também não aprofundou a discussão de dois outros aspectos: questões de validade, tais como validade de conclusão, interna, construto e externa (Trochim & Donnelly, 2008) e questões éticas. Quanto ao primeiro aspecto, entendo que a validade de um estudo passa por seu apropriado desenho e execução. Por exemplo, o cuidado em vincular definições conceituais e operacionais dos fatores contribui para a validade de construto; a aleatorização de sujeitos às condições experimentais contribui para a validade interna; a seleção de sujeitos heterogêneos contribui para a validade externa; e a decisão do número de participantes influencia a validade de conclusão do estudo. Ainda que tenha procurado vincular a discussão de aspectos específicos explorados neste artigo aos seus potenciais efeitos de validação, esse não era o propósito principal e foi feito apenas quando e aonde pareceu conveniente. Entretanto, destaco que o fato de uma ameaça à validade ser possível em determinado estudo, não significa que ela seja plausível; para verificar a plausibilidade de uma ameaça à validade de um estudo é preciso se assegurar que essa ameaça é responsável pelo efeito observado, e o tamanho desse efeito (Shadish et al., 2002). Se uma ameaça provoca apenas um efeito pequeno que não é capaz de explicar o tamanho do efeito encontrado em um estudo ou se essa ameaça provoca um efeito na direção contrária do resultado obtido no estudo, não é possível afirmar que essa seja uma ameaça plausível (Shadish et al., 2002). É importante considerar esse critério de plausibilidade na avaliação crítica da validade dos resultados de um estudo experimental.

Quanto ao segundo aspecto, são diversas as questões éticas associadas ao uso de experimentos em pesquisa social (ver p. ex., Babbie, 2010), com destaque para: quão voluntária é a participação, em particular quando os participantes são estudantes de graduação; eventuais prejuízos (p. ex., emocionais) causados aos participantes em razão da pesquisa; garantia de anonimato e confidencialidade; e qual e em que extensão fornecer aos eventuais participantes informação sobre o estudo e sobre o pesquisador antes da obtenção de concordância em participar, com o propósito de evitar que os participantes se sintam 'enganados' (*deception*). Em particular, esse último aspecto tem sido tema de ampla discussão no ambiente acadêmico (ver p. ex., Hertwig e Ortmann, 2008). Nesse sentido, entende-se como fundamental a constituição de comitês de ética para os quais pesquisadores possam submeter seus protocolos de pesquisa para avaliação e aprovação. Importante ainda a obtenção de consentimento dos participantes uma vez sejam informados os riscos e procedimentos envolvidos no estudo (Trochim & Donnelly, 2008).

Por fim, considerando a dificuldade de acesso a respondentes para pesquisas de levantamento ou estudos qualitativos, a limitação de bases de dados adequadas para determinados propósitos de pesquisa, além da própria natureza da questão de pesquisa que dificulta o uso de determinados métodos de pesquisa, tal como indicado em alguns exemplos apresentados neste artigo, entende-se que o uso de pesquisa experimental pode ser uma importante alternativa de pesquisa que pesquisadores brasileiros poderiam mais amplamente explorar, o que, de fato, parece já estar acontecendo. Aspectos regulatórios específicos do contexto brasileiro que afetam a prática contábil em organizações e outros agentes, incluindo investidores e gestores, podem representar oportunidades de pesquisa relevantes a serem conduzidas por meio de estudos experimentais. Um exemplo é o processo de harmonização das normas contábeis em que um evento identificável afetou um número representativo de empresas e cujos efeitos poderiam ser explorados por meio de desenhos quase-experimentais, tal como em Iliev (2010), em relação à passagem da lei Sarbanes-Oxley no contexto norte-americano. Ainda, estudos experimentais ou quase-experimentais podem representar uma importante forma de aproximação entre pesquisadores e organizações atuantes no Brasil por meio de verificação do efeito de diferentes alternativas de elementos de controle gerencial sobre variáveis de interesse. Por exemplo, uma organização com duas divisões pode não está segura quanto a implementar ou não um sistema de avaliação de desempenho baseado em custo padrão. Um estudo experimental poderia permitir a essa empresa verificar potenciais efeitos dessa implementação em uma divisão, enquanto mantendo o sistema atual na outra divisão, antes de decidir quanto a uma total implementação. Por esses motivos é que entendo que o potencial de utilização do método experimental na pesquisa contábil brasileira esteja apenas em seu início, o que faz com que sejam bem-vindos, em programas de pós-graduação da área, workshops, dentre outros, treinamentos específicos sobre o uso apropriado desse método de pesquisa de modo a contribuir para a validade das inferências decorrentes de estudos experimentais em contabilidade desenvolvidos no Brasil.

## REFERÊNCIAS

- Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *The Review of Economic Studies*, 76(2), 451-469.
- Babbie, E. (2010). *The Practice of Social Research*, 12a ed. California, Wadsworth Cengage Learning.
- Bloomfield, R. J., & Wilks, T. J. (2000). Disclosure effects in the laboratory: Liquidity, depth, and the cost of capital. *The Accounting Review*, 75(1), 13-41.
- Bol, J. C., & Smith, S. D. (2011). Spillover effects in subjective performance evaluation: Bias and the asymmetric influence of controllability. *The Accounting Review*, 86(4), 1213-1230.
- Bonner, S. E., Hastie, R., Sprinkle, G. B., & Young, S. M. (2000). A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research*, 12(1), 19-64.
- Bryant, S. M., Hunton, J. E., & Stone, D. N. (2004). Internet-based experiments: Prospects and possibilities for behavioral accounting research. *Behavioral Research in Accounting*, 16(1), 107-129.
- Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of risk and uncertainty*, 19(1-3), 7-42.

Carpenter, J., Harrison, G., & List, J. A. (2005). Field experiments in economics: An introduction. In: *Field experiments in economics*. Published online: 09 Mar 2015, 1-15.

Chen, C. X., Pesch, H. L., Wang, L. W. (2016). The Effect of Pay Level on Employee Selection and Performance in Mission-Driven Organizations. Working Paper.

Choi, J., Hecht, G. W., Tafkov, I., & Towry, K. L. (2015). Vicarious Learning Under Implicit Contracts. *The Accounting Review*, 91(4), 1087-1108.

Christ, M. H., Sedatole, K. L., & Towry, K. L. (2012). Sticks and carrots: The effect of contract frame on effort in incomplete contracts. *The Accounting Review*, 87(6), 1913-1938.

Christ, M. H.; Emmett, S. A.; Tayler, W. B.; & Wood, D. A. (2016). Compensation or feedback: Motivating performance in multidimensional tasks. *Accounting, Organizations and Society*, 50, 27-40.

Christ, M. H., Vance, T. (2015). Cascading Controls: The Effects of Manager Incentive Frame on Subordinate Behavior. Working Paper.

Cloyd, C. B., Pratt, J., & Stock, T. (1996). The use of financial accounting choice to support aggressive tax positions: Public and private firms. *Journal of Accounting research*, 34(1), 23-43.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18(1), 105-115.

Elliott, W. B., Hodge, F. D., Kennedy, J. J., & Pronk, M. (2007). Are MBA students a good proxy for nonprofessional investors? *The Accounting Review*, 82(1), 139-168.

Festinger, L. (1953). Laboratory experiments. In: Festinger, L. & Katz, D. *Research methods in the behavioral sciences*. New York, The Dryden Press, 136-172.

French, J. R. (1953). Experiments in field settings. In: Festinger, L. & Katz, D. *Research methods in the behavioral sciences*. New York, The Dryden Press, 98-135.

Frezatti, F., de Aguiar, A. B., de Araujo Wanderley, C., & Malagueño, R. (2015). A pesquisa em contabilidade gerencial no brasil: desenvolvimento, dificuldades e oportunidades. *Revista Universo Contábil*, 11(1), 47.

Friedman, D., & Sunder, S. (1994). *Experimental methods: a primer for economists*. Cambridge, Cambridge University Press, 1994.

Gall, M. D., Borg, W. R., & Gall, J. P. (2007). *An introduction to educational research*, 8a ed., Pearson

Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly journal of economics*, 115(3), 791-810.

Gormley, T. A., Matsa, D. A., & Milbourn, T. (2013). CEO compensation and corporate risk: Evidence from a natural experiment. *Journal of Accounting and Economics*, 56(2), 79-101.

Guo, L., Libby, T., & Liu, X. K. (2017). The effects of vertical pay dispersion: Experimental evidence in a budget setting. *Contemporary Accounting Research*, 34(1), 555-576.

Hannan, R. L., Towry, K. L., & Zhang, Y. M. (2013). Turning up the volume: An experimental investigation of the role of mutual monitoring in tournaments. *Contemporary Accounting Research*, 30(4), 1401-1426.

Hannan, R. L., Rankin, F. W., & Towry, K. L. (2010). Flattening the organization: The effect of organizational reporting structure on budgeting effectiveness. *Review of Accounting Studies*, 15(3), 503-536.

Harrison, G. W. (2005). Field experiments and control. In: *Field experiments in economics*. Published online: 09 Mar 2015, 17-50.

Hecht, G., Tafkov, I., & Towry, K. L. (2012). Performance spillover in a multitask environment. *Contemporary Accounting Research*, 29(2), 563-589.

Hertwig, R., & Ortmann, A. (2008). Deception in experiments: Revisiting the arguments in its defense. *Ethics & Behavior*, 18(1), 59-92.

Hesford, J. W., Lee, S. H. S., Van der Stede, W. A., & Young, S. M. (2006). Management accounting: a bibliographic study. In: Chapman, C. S., Hopwood, A. G., & Shields, M. D. (Eds.), *Handbooks of management accounting research*, v. 1, Oxford: Elsevier, 3-26.

Homero Jr., P. F. (2016). Crítica Metodológica e Epistemológica de Pesquisas Contábeis Experimentais Publicadas no Brasil. *Revista de Educação e Pesquisa em Contabilidade (REPeC)*, 10(2), 220-233.

Iliev, P. (2010). The Effect of SOX Section 404: Costs, Earnings Quality, and Stock Prices. *Journal of Finance*, 65(3), 1163-1196.

Kachelmeier, S. J., Thornock, T. A., & Williamson, M. G. (2016). Communicated values as informal controls: Promoting quality while undermining productivity?. *Contemporary Accounting Research*, 33(4), 1411-1434.

Kelly, K. O. (2007). Feedback and incentives on nonfinancial value drivers: Effects on managerial decision making. *Contemporary Accounting Research*, 24(2), 523-556.

Koonce, L., Williamson, M. G., & Winchel, J. (2010). Consensus information and nonprofessional investors' reaction to the revelation of estimate inaccuracies. *The Accounting Review*, 85(3), 979-1000.

Kuang, X. J., & Moser, D. V. (2011). Wage Negotiation, Employee Effort, and Firm Profit under Output-Based versus Fixed-Wage Incentive Contracts. *Contemporary Accounting Research*, 28(2), 616-642.

Leary, M. R. (2012). *Introduction to behavioral research methods*, 6a ed. New Jersey, Pearson Education Inc.

Libby, R., Bloomfield, R., & Nelson, M. W. (2002). Experimental research in financial accounting. *Accounting, Organizations and Society*, 27(8), 775-810.

Libby, T., Salterio, S. E., & Webb, A. (2004). The balanced scorecard: The effects of assurance and process accountability on managerial judgment. *The Accounting Review*, 79(4), 1075-1094.

Lourenço, S. M. (2015). Monetary Incentives, Feedback, and Recognition-Complements or Substitutes? Evidence from a Field Experiment in a Retail Services Company. *The Accounting Review*, 91(1), 279-297.

Maas, V. S., Van Rinsum, M., & Towry, K. L. (2012). In search of informed discretion: An experimental investigation of fairness and trust reciprocity. *The Accounting Review*, 87(2), 617-644.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: a model comparison perspective*. California, Wadsworth Publishing Company.

Myers, J. L. (1979). *Fundamentals of experimental design*, 3 ed. Boston, Allyn and Bacon, Inc.

Ortmann, A. (2005). Field experiments in economics: Some methodological caveats. In: Field experiments in economics. Published online: 09 Mar 2015, 51-70.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5(5), 411-419.

Peecher, M. E., & Solomon, I. (2001). Theory and experimentation in studies of audit judgments and decisions: Avoiding common research traps. *International Journal of Auditing*, 5(3), 193-203.

Pinsker, R. (2011). Primacy or recency? A study of order effects when nonprofessional investors are provided a long series of disclosures. *Behavioral Research in Accounting*, 23(1), 161-183.

Presslee, A., Vance, T. W., & Webb, R. A. (2013). The effects of reward type on employee goal setting, goal commitment, and performance. *The Accounting Review*, 88(5), 1805-1831.

Reips, U. D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In: Birnbaum, M. H. (ed.), *Psychological experiments on the Internet*, California: Academic Press, 89-117.

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

Smith, E. R. Research design. Reis, H. T., & Judd, C. M. (2000). *Handbook of research methods in social and personality psychology*. Cambridge University Press. 27-48.

Sprinkle, G. B. (2003). Perspectives on experimental research in managerial accounting. *Accounting, Organizations and Society*, 28(2), 287-318.

Tan, H. T., Wang, E. Y., & Zhou, B. (2015). How does readability influence investors' judgments? Consistency of benchmark performance matters. *The Accounting Review*, 90(1), 371-393.

Trochim, William M.K. and James P. Donnelly. *The Research Methods Knowledge Base*, 3a. ed. Ohio, Cengage Learning, 2008.