RASI REVISTA DE ADMINISTRAÇÃO, SOCIEDADE E INOVAÇÃO

Revista de Administração, Sociedade e Inovação - v. 12, n. 4, pp. 01-18, 2026 – www.rasi.vr.uff.br

Diretrizes para a elaboração de questionários: do planejamento à análise dos dados

Ualison Rébula de Oliveira – Universidade Federal Fluminense – ualisonrebula@id.uff.br
Marco Antônio Conejero – Universidade Federal Fluminense – marcoac@id.uff.br

∴

Editorial

https://doi.org/10.20401/rasi.12.4.1410

RASI REVISTA DE ADMINISTRAÇÃO, SOCIEDADE E INOVAÇÃO

Revista de Administração, Sociedade e Inovação - v. 12, n. 4, pp. 01-18, 2026 – www.rasi.vr.uff.br

Diretrizes para a elaboração de questionários: do planejamento à análise dos dados

Esta série de editoriais nasceu de uma constatação feita em um levantamento (*survey*) com os alunos do nosso Programa de Pós-Graduação em Administração (PPGA/UFF Volta Redonda), segundo os quais o capítulo de Materiais e Métodos está entre os mais difíceis de escrever em uma dissertação. Por isso, desde o último número de 2023, nossos editoriais se dirigem a um leitor específico: alunos de programas de mestrado e doutorado em Administração. O conteúdo pode ser útil a outras áreas do conhecimento, mas foi pensado para orientar esses alunos. Este é o décimo editorial da série e, em um número especial, trata do instrumento de coleta de dados, tema invariavelmente presente nas pesquisas que eles conduzem.

A escolha tem uma razão prática. É no questionário, no roteiro ou no formulário que o projeto de pesquisa se converte em dado. Um item mal redigido, uma escala sem rótulos ou uma sequência de perguntas que induz respostas produzem "dados com defeito", que não correspondem ao que o pesquisador pretendia medir. Nenhuma técnica de análise aplicada depois consegue separar o que o respondente pensa, sente ou julga daquilo que o instrumento o levou a responder. Entre esses instrumentos, este editorial se concentra no questionário com itens fechados e abertos, com atenção especial às pesquisas que medem a mudança provocada por uma intervenção. Temas como a adaptação de escalas já validadas, a verificação da confiabilidade, a elaboração de roteiros de entrevista, a amostragem e os cuidados éticos ficam fora desse recorte e poderão ser tratados em outros editoriais da série.

Nas bancas de qualificação de que participamos, já encontramos instrumentos com perguntas que não contribuíam para os objetivos da pesquisa (o famoso "encher linguiça") e, no sentido oposto, instrumentos sem as perguntas necessárias para responder ao problema investigado. A situação mais frequente, porém, era a do aluno que ainda não havia pensado em como trataria os dados. Houve também quem propusesse técnicas inadequadas para analisá-los. Felizmente, esses problemas apareceram na qualificação, quando ainda havia tempo para corrigir o instrumento antes da coleta.

Para este editorial, selecionamos nove artigos seminais que, em conjunto, cobrem o percurso do instrumento da concepção à análise. Todos foram publicados em periódicos classificados no primeiro quartil (Q1). Em consulta ao Google Acadêmico feita em novembro de 2026, oito deles superavam mil citações, e o de Braun e Clarke (2006) passava de 360 mil. Trabalhamos somente com essas obras para que o leitor possa conferir na fonte cada recomendação feita aqui.

Cada artigo trata de uma parte desse percurso. Tourangeau e Rasinski (1988) descrevem como o respondente processa uma pergunta. Podsakoff et al. (2003) organizam as fontes de viés que decorrem do próprio método de coleta. Howard e Dailey (1979) mostram o que pode distorcer o resultado quando o pesquisador pede ao respondente que avalie a si mesmo antes e depois de uma intervenção, como um curso ou treinamento, para medir o quanto ele mudou. Beatty e Willis (2007) revisam como os pesquisadores testam um questionário antes da coleta por meio da entrevista cognitiva, na qual pequenos grupos de participantes respondem aos itens e relatam como os interpretaram. Jick (1979) e Ivankova et al. (2006) discutem a combinação de dados quantitativos e qualitativos. Jamieson (2004), Conover e Iman (1981) e Braun e Clarke (2006) orientam a análise dos itens fechados e abertos.

Os exemplos de problemas discutidos ao longo deste editorial foram retirados desses nove artigos, que os apresentam a partir de estudos empíricos ou de casos ilustrativos. Para

mostrar soluções aplicadas, recorremos também a trechos de um instrumento que elaboramos para avaliar, com doutorandos, um workshop sobre análise documental. Antes de passar a esse conteúdo, registramos que este número (Volume 12, Número 4) reúne seis trabalhos, cinco artigos e um caso de ensino. Os artigos tratam das metas de resíduos sólidos em dois municípios do Sul Fluminense, das diretrizes internacionais de impacto de tráfego sob a ótica da bicicleta, da inovação orientada à sustentabilidade em *clusters* industriais, do papel das tecnologias nas estratégias de serviços e do uso de dados na Defensoria Pública do Estado do Amazonas. O caso de ensino acompanha os desafios de um executivo na reestruturação de uma franquia. Todos são apresentados ao final deste editorial.

O instrumento como tecnologia de medição

É comum que o aluno trate o questionário como uma lista de perguntas sobre o tema da pesquisa. Essa visão deixa de lado duas coisas. A primeira é o que acontece entre a leitura do item (cada pergunta ou afirmação a que o respondente reage) e a marcação da resposta, intervalo em que se forma boa parte do erro de medida, isto é, da diferença entre o que o pesquisador pretendia medir e o que a resposta de fato registra. A segunda é a análise, pois um item redigido sem pensar em como seus dados serão tratados pode tornar a análise difícil ou até inviável, algo que costuma pegar o aluno de surpresa depois da coleta, quando já não há como corrigir o instrumento.

Tourangeau e Rasinski (1988) descrevem a resposta a uma pergunta de atitude como um processo em quatro estágios. Primeiro, o respondente interpreta a pergunta e identifica de que assunto ela trata. Em seguida, recupera da memória crenças e sentimentos ligados ao assunto. Depois, usa esse material para formar um julgamento. Por fim, ajusta o julgamento às opções de resposta oferecidas e, antes de marcar, pode alterar a resposta para que pareça coerente com as anteriores ou socialmente aceitável. A resposta registrada pode ser, assim, um meio-termo entre o que o respondente de fato pensa e a pressão por coerência ou por aprovação social. Os autores observam que, quando o interesse do respondente é baixo e o tempo é curto, o processo tende a ser superficial, e a memória entrega o que está mais acessível, e não necessariamente o que é mais importante para quem responde.

Cada um desses estágios pode ser afetado pelos itens anteriores. Tourangeau e Rasinski (1988) relatam um estudo em que perguntaram a opinião dos respondentes sobre um projeto de lei pouco conhecido. Quando um bloco de perguntas sobre inflação vinha imediatamente antes, 27,5% se declaravam favoráveis ao projeto e 55,0% diziam não saber. Com perguntas neutras antes, esses percentuais eram de 12,5% e 62,5%. As perguntas anteriores deram aos respondentes uma interpretação para um tema que eles desconheciam. Em uma replicação com outro tema pouco conhecido, o efeito desapareceu quando a redação do item esclarecia o assunto, o que sugere definir no próprio item aquilo que o respondente pode não conhecer.

Podsakoff et al. (2003) ampliam esse quadro. Os autores lembram que o erro de medida tem um componente aleatório e um sistemático, e que o segundo é mais grave, porque oferece para os resultados uma explicação concorrente, independente da que o pesquisador propôs. Em outras palavras, o que se atribui ao fenômeno pode ter vindo do método com que ele foi medido. Entre as principais fontes desse erro sistemático estão os vieses de método, que decorrem da forma de coletar os dados, e não do fenômeno estudado. Esses vieses não atuam em uma direção única. Podem inflar ou reduzir a relação observada entre os construtos, isto é, entre os conceitos

que o instrumento pretende medir, e levar tanto a afirmar uma relação que não existe quanto a deixar de detectar uma que existe. Os autores os agrupam em quatro conjuntos:

i) efeitos do respondente comum, que surgem porque é a mesma pessoa quem responde a todas as perguntas do instrumento e leva para cada resposta as próprias tendências, como o desejo de parecer coerente, as crenças sobre o que costuma andar junto (teorias implícitas), a vontade de dar a resposta socialmente aprovada (desejabilidade social), a tendência a avaliar melhor quem se conhece e de quem se gosta (leniência), a tendência a concordar com os itens independentemente do conteúdo (aquiescência) e o estado de humor em que o respondente se encontra ao responder;

ii) efeitos das características do item, como a desejabilidade social embutida na própria redação, as pistas sobre a resposta esperada, a ambiguidade, a redação na forma afirmativa ou negativa e o uso dos mesmos formatos e rótulos de escala (âncoras) em todo o instrumento;

iii) efeitos do contexto do item, como a influência dos itens anteriores sobre os seguintes, o tamanho das escalas (quanto menos itens elas têm, mais as respostas anteriores permanecem acessíveis na memória e influenciam as seguintes), a colocação de itens neutros em meio a itens positivos ou negativos, que os faz assumir o tom desses itens, e a forma como itens de construtos diferentes são intercalados ou agrupados no questionário;

iv) efeitos do contexto de medição, quando construtos diferentes são medidos no mesmo momento, no mesmo local e pelo mesmo meio (por exemplo, no mesmo questionário on-line), o que pode criar entre eles uma relação artificial, que não decorre do conteúdo dos construtos.

A preocupação central dos autores é com pesquisas que relacionam construtos medidos pela mesma pessoa, como o comportamento do líder e a satisfação dos funcionários, exemplo discutido no próprio artigo, porque nelas o viés pode distorcer a relação observada. Os mecanismos descritos, porém, atuam sobre qualquer item, e cada decisão de redação, de escala, de ordem e de aplicação interfere no dado. Nem todos os remédios propostos pelos autores cabem na realidade de uma dissertação ou tese, mas boa parte deles depende apenas de cuidado na elaboração do instrumento. Por isso o capítulo de método precisa mostrar que essas decisões foram tomadas de propósito.

Roteiro de elaboração do instrumento

O Quadro 1 reúne em dez etapas o percurso de elaboração do instrumento. O roteiro é uma síntese que fizemos a partir das nove obras, e não a proposta de um autor específico. Para cada etapa, o quadro indica o que fazer, em que autores a recomendação se apoia e o que registrar no texto do método, ou em apêndice, para mostrar que a etapa foi cumprida. Assim, o aluno pode consultá-lo a cada decisão ao elaborar o questionário e, ao final, usá-lo como guia para redigir o método.

As etapas formam quatro blocos. As cinco primeiras compõem a concepção do instrumento, que começa pelo planejamento (etapas 1 e 2), em que se define o que será medido, como será analisado e em que desenho de pesquisa, e segue com a construção propriamente dita (etapas 3 a 5), que trata da redação, da escala e da sequência dos itens. O segundo bloco põe o instrumento à prova no pré-teste (etapa 6), e o terceiro cuida da aplicação (etapa 7). O quarto bloco, formado pelas três últimas etapas, antecipa a análise dos itens fechados, dos abertos e da integração entre eles. A ordem tem um motivo. Começar pela redação e deixar a análise para depois está na origem de boa parte dos problemas discutidos na seção seguinte.

Quadro 1. Roteiro de elaboração do instrumento de coleta de dados

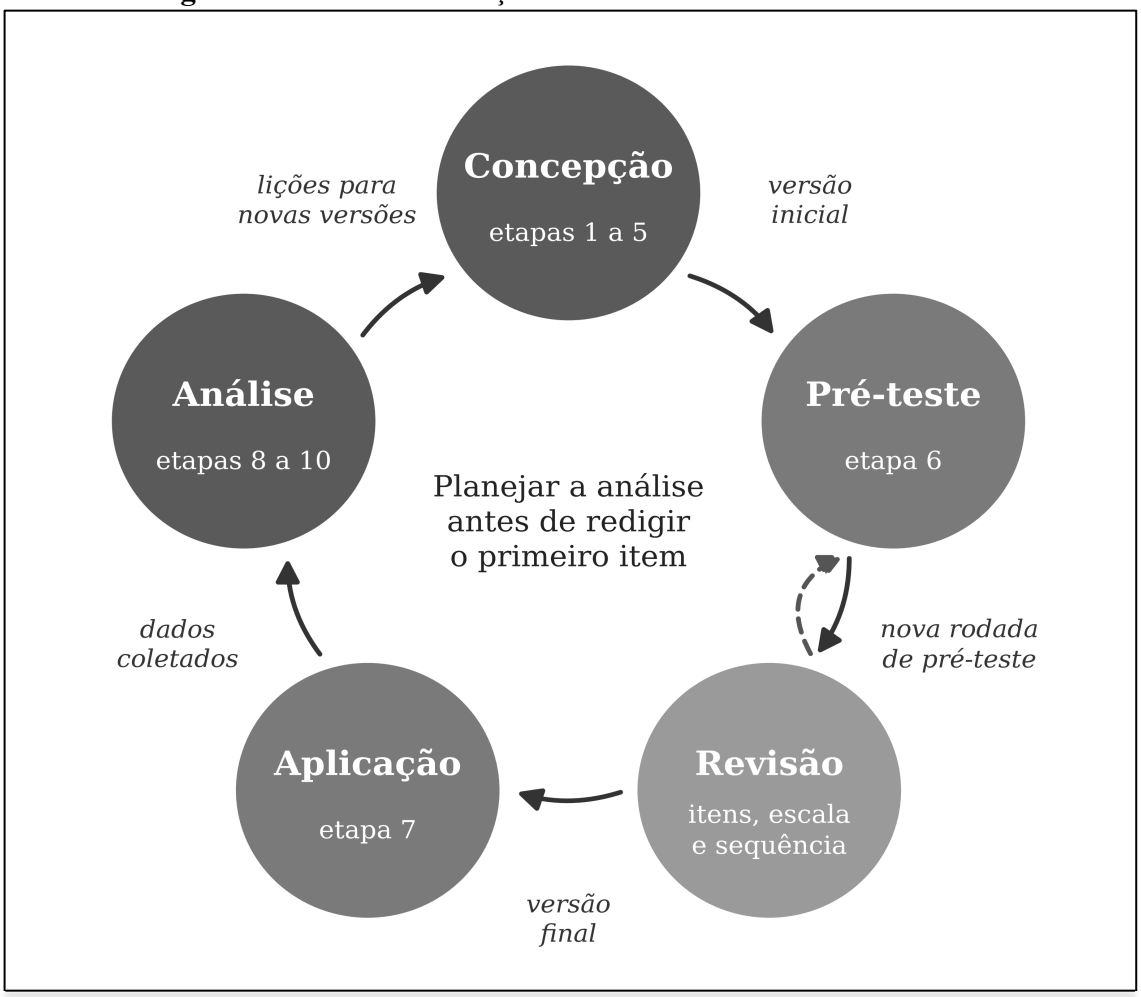
Etapa	O que fazer	Base na literatura	O que registrar no método
1. Definir o que será medido e como será analisado	Definir com precisão o conceito a ser medido (o construto) e decidir, antes de escrever qualquer item, que tipo de dado cada pergunta vai gerar e qual análise será feita com ele.	Jamieson (2004); Jick (1979)	Quadro que liga cada item a um objetivo da pesquisa e à análise prevista.
2. Escolher o desenho de coleta	Decidir quando e quantas vezes medir. Para medir mudança pela autoavaliação, pedir ao final da intervenção duas respostas para cada item, uma sobre como o respondente está agora e outra sobre como se via antes, e, se possível, manter a medida feita no início. Decidir se as perguntas fechadas e abertas serão aplicadas juntas ou em fases separadas.	Howard e Dailey (1979); Ivankova et al. (2006); Jick (1979)	Quando e quantas vezes se mediu, e por quê; havendo fase qualitativa posterior, o ponto em que ela se conecta à quantitativa.
3. Redigir os itens	Pedir um único julgamento por item, explicar no próprio item os termos que o respondente pode não conhecer, usar frases simples e o vocabulário do público pesquisado.	Podsakoff et al. (2003); Tourangeau e Rasinski (1988)	A versão inicial dos itens, que será levada ao pré-teste.
4. Definir a escala e as instruções	Dar nome a todos os pontos da escala, inclusive ao do meio (por exemplo, “nem concordo nem discordo”); evitar valores numéricos bipolares (de −3 a +3), salvo quando o construto for bipolar por natureza; usar com cautela itens redigidos na forma negativa; registrar que a escala indica ordem, mas não distâncias iguais entre os pontos (escala ordinal).	Podsakoff et al. (2003); Tourangeau e Rasinski (1988); Jamieson (2004)	A escala adotada, com todos os rótulos, e a justificativa do nível de mensuração assumido.
5. Organizar a sequência	Verificar se ler um item pode mudar como o respondente entende ou responde o seguinte. Esse efeito também pode ser usado a favor, com perguntas de aquecimento que ativem a memória do que se quer medir. Contrabalançar as ordens, sabendo que isso pode quebrar a lógica do questionário, ou manter ordem fixa, do geral para o específico, testando-a no pré-teste.	Tourangeau e Rasinski (1988); Podsakoff et al. (2003)	A ordem adotada e a razão dela; havendo contrabalanceamento, como as versões foram distribuídas.
6. Pré-testar	Antes da coleta, aplicar o instrumento a um pequeno grupo e investigar como entenderam cada item, pedindo que pensem em voz alta enquanto respondem ou fazendo perguntas sobre o que entenderam (entrevista cognitiva); trabalhar em rodadas de 5 a 15 entrevistas, revisando os itens entre uma rodada e outra, com um critério para encerrar as rodadas, como a ausência de problemas novos.	Beatty e Willis (2007)	Número de rodadas e de participantes, problemas encontrados, alterações feitas entre uma rodada e outra e a versão final do instrumento, em apêndice.
7. Preparar as condições de aplicação	Garantir o anonimato; dizer ao respondente que não há respostas certas ou erradas; quando possível, coletar partes do instrumento em momentos, locais ou meios diferentes, ou com respondentes diferentes; se as respostas do início e do fim da intervenção serão comparadas, pedir que cada respondente crie um código pessoal (por exemplo, as iniciais da mãe e o dia do próprio aniversário) e o repita nas duas aplicações, o que permite ligá-las sem identificá-lo.	Podsakoff et al. (2003); Howard e Dailey (1979)	Meio, local e responsável pela aplicação, a forma de garantir o anonimato e, se houver, o código usado para ligar as respostas.
8. Analisar os itens fechados	Descrever os resultados pela mediana (o valor central), pela moda (a resposta mais frequente) e pela distribuição das respostas; usar testes que trabalham com a ordem dos valores (testes não paramétricos), como o de Wilcoxon; informar como foram tratados os valores repetidos (empates) e as diferenças iguais a zero.	Jamieson (2004); Conover e Iman (1981)	Estatísticas escolhidas e sua justificativa; tratamento dado aos empates e às diferenças nulas.

9. Analisar os itens abertos	Seguir as seis fases da análise temática e declarar as escolhas feitas, por exemplo se os temas partem dos dados ou de uma teoria e se a análise fica no que foi dito ou busca os significados por trás das respostas.	Braun e Clarke (2006)	As escolhas declaradas (se os temas partem dos dados ou de uma teoria; se a análise fica no que foi dito ou busca o que está por trás) e a lista de códigos usada.
10. Integrar os resultados	Comparar o que dizem os números e as respostas abertas e explicar as divergências, em vez de ignorá-las. Se a parte qualitativa vier depois da quantitativa (desenho sequencial), planejá-la a partir dos resultados da primeira fase.	Jick (1979); Ivankova et al. (2006)	O ponto em que as duas fontes foram confrontadas e como as divergências foram tratadas.

Fonte: Elaboração própria, a partir dos autores citados (2026).

A Figura 1 mostra que esse percurso não é linear. Depois do pré-teste, o pesquisador revisa os itens, a escala e a sequência e, se necessário, faz uma nova rodada de entrevistas antes de aplicar o instrumento. Do mesmo modo, o que se aprende na análise orienta a versão seguinte, seja em uma nova coleta, seja em pesquisas futuras que usem o mesmo instrumento.

Figura 1. Ciclo de elaboração do instrumento de coleta de dados



Fonte: Elaboração própria (2026).

Na concepção do instrumento, as etapas 2 a 5 concentram decisões que o aluno costuma tomar sem perceber suas consequências, e por isso as comentamos com mais detalhe.

A primeira é a escolha do desenho de coleta quando a pesquisa pretende medir mudança, por exemplo, o quanto um curso melhorou certa habilidade dos participantes na avaliação deles próprios. Howard e Dailey (1979) estudaram essa situação com 21 profissionais responsáveis por entrevistas de seleção, que participaram de uma oficina de 40 horas. No início da oficina, cada participante avaliou a própria habilidade em uma escala (pré-teste). No encerramento, respondeu duas vezes a cada item, primeiro sobre como se percebia naquele momento (pós-teste) e logo em seguida sobre como se percebia antes da oficina (avaliação retrospectiva). Para saber qual comparação refletia melhor a mudança, os autores pediram a avaliadores treinados, que não participaram da oficina e são chamados de juízes na literatura, que analisassem gravações de entrevistas feitas pelos participantes no início e no fim da oficina.

A mudança medida pela comparação entre pré-teste e pós-teste praticamente não teve relação com a mudança observada pelos avaliadores, com correlação média de −0,06, próxima de zero. Já a mudança medida pela comparação entre avaliação retrospectiva e pós-teste acompanhou bem mais a avaliação deles, com correlação média de 0,43. Em outras palavras, perguntar ao participante, no fim da oficina, como ele se via antes produziu uma medida de mudança mais próxima da avaliação externa do que usar a resposta que ele deu no primeiro dia.

Os autores atribuem a diferença a uma mudança no critério com que o participante se avalia e ilustram o problema com uma participante que, antes de uma oficina sobre liderança, se considera uma líder mediana. A oficina amplia sua compreensão do que é liderar, e ela percebe que, antes, estava abaixo da média. Se ela melhorar e chegar à média, marcará “mediana” no pré-teste e no pós-teste, e o pesquisador concluirá, de forma equivocada, que a oficina não teve efeito. Na avaliação retrospectiva, feita no fim e já com o novo critério, ela marcaria “abaixo da média” para o antes e “mediana” para o depois, e a melhora apareceria.

Para mostrar como essas recomendações se traduzem em um instrumento real, recorremos a um workshop metodológico de oito horas que realizamos com doutorandos da turma de 2026 do PPGA/UFF Volta Redonda, na disciplina Procedimentos e Métodos de Pesquisa em Administração. Nele, apresentamos um protocolo de análise documental organizado em cinco fases. Os participantes estudaram sua fundamentação teórica, discutiram-no a partir de casos de ensino e o aplicaram na prática. Ao final, responderam a um questionário em quatro blocos, que tratavam da competência autoavaliada antes e depois do workshop, da usabilidade e da aplicabilidade do protocolo, do papel dos casos de ensino e da prática na aprendizagem e da governança dos dados e da validação do protocolo. É desse questionário que vêm os trechos usados como exemplo ao longo do texto.

Esse instrumento adotou a avaliação retrospectiva. Ao final do workshop, cada afirmação aparecia em par, uma sobre o antes e outra sobre o depois, e os dois itens mantinham o mesmo conteúdo, mudando apenas a referência ao momento avaliado:

> *Antes do workshop, eu me sentia seguro(a) para conduzir uma Análise Documental completa na minha própria tese.*
> *Depois do workshop, sinto-me seguro(a) para conduzir uma Análise Documental completa na minha própria tese.*

Howard e Dailey (1979) lembram que duas medidas só podem ser comparadas se forem feitas com a mesma régua, o que os autores chamam de métrica comum. No par de itens, a

avaliação retrospectiva favorece essa condição, porque o respondente avalia o antes e o depois no mesmo momento e com o mesmo critério. A redação simétrica completa esse cuidado. Se o texto mudasse de um item para o outro, a diferença entre as respostas poderia refletir a mudança de redação, e não a do respondente. Além disso, cada item pede um único julgamento, sobre uma situação concreta.

A segunda é a redação dos itens. Beatty e Willis (2007) relatam três casos, identificados em entrevistas cognitivas, que mostram como defeitos de redação escapam a quem escreve o item. O primeiro item perguntava:

> *Nos últimos 12 meses, quantas vezes você foi atendido ou conversou por telefone sobre sua saúde física, emocional ou mental com um médico de família ou clínico geral? (Beatty & Willis, 2007, p. 297, tradução nossa)*

Um participante respondeu "zero", embora outras respostas indicassem atendimento recente. Quando o entrevistador fez novas perguntas para entender a resposta, o participante explicou que tinha entendido a pergunta como restrita às conversas por telefone. A referência ao atendimento havia se perdido na extensão do enunciado, e o erro se mostrou comum nas entrevistas seguintes.

O segundo item pedia o número de dias, nos últimos 30, em que a saúde física do respondente não esteve boa. Vários participantes não conseguiram responder com um número, porque, como disse um deles, um dia é em parte bom e em parte ruim. O item supunha que o dia fosse uma unidade adequada para descrever a saúde, o que não vale para quem tem dias de qualidade variável.

O terceiro perguntava se o respondente estava "atualmente em tratamento" para artrite. A pergunta era simples para quem tinha artrite grave ou não tinha a doença, mas gerava dúvida para quem havia consultado um médico pela última vez há mais de um ano. Especificar o termo resolveria a dúvida desse grupo, mas tornaria o item mais longo, mais cansativo e talvez mais confuso para os demais. Por isso os autores concluem que a entrevista cognitiva serve menos para encontrar a pergunta perfeita do que para fundamentar as decisões de redação.

Os três casos ilustram recomendações que Podsakoff et al. (2003) reúnem a partir da literatura sobre pesquisas de levantamento (surveys). Entre elas estão definir termos ambíguos ou pouco familiares, dar exemplos quando o conceito for vago, manter os itens simples, específicos e concisos, evitar sintaxe complicada e decompor em itens separados toda pergunta que envolva mais de uma possibilidade. O instrumento do workshop aplicou a primeira delas, a definição de termos pouco familiares, nos itens em que o respondente avaliava a utilidade de cada ferramenta do protocolo. Como o nome da ferramenta, sozinho, não diria ao respondente o que ela é, cada item descrevia a ferramenta. Em vez de pedir a avaliação da "matriz de extração", por exemplo, o instrumento apresentava o item assim:

> ENUNCIADO: *Para cada uma das ferramentas e técnicas operacionais abaixo, avalie a sua utilidade prática para estruturar, orientar e operacionalizar a condução da Análise Documental.*
>
> ITEM: *Matriz de extração de dados em planilha, com colunas para identificador do documento, autor, data, trecho de evidência empírica e categoria teórica.*
>
> ESCALA: *1 (Nada útil) a 5 (Extremamente útil).*

Um respondente que não lembrasse o nome da ferramenta ainda saberia o que estava avaliando, o que reduz o espaço para interpretações próprias e a dependência do contexto para interpretar o item.

A terceira é a escala. Podsakoff et al. (2003) recomendam dar nome verbal ao ponto do meio (por exemplo, “nem concordo nem discordo”) e evitar escalas numéricas bipolares, do tipo −3 a +3, salvo quando o construto for bipolar por natureza, duas medidas que podem reduzir a tendência a concordar com tudo (aquiescência). Tourangeau e Rasinski (1988) acrescentam que rotular todas as opções pode diminuir a influência de pontos de referência alheios à opinião do respondente, como os itens anteriores (efeitos de ancoragem), mas advertem que isso também pode transformar o próprio ponto do meio em âncora e atrair as respostas para o centro. Os itens invertidos, redigidos na forma negativa para obrigar o respondente a ler com atenção, também exigem cautela. Segundo Podsakoff et al. (2003), esses itens podem formar fatores artificiais, compostos apenas pelos itens negativos, mesmo quando só 10% dos respondentes deixam de perceber a inversão, e esses fatores podem desaparecer quando os mesmos itens são reescritos na forma positiva. Os autores apontam ainda o uso dos mesmos formatos e rótulos de escala (âncoras) em todo o instrumento como fonte de viés, mas advertem que alterá-los pode mudar o sentido do construto, e pedem que não se sacrifique a validade da escala para reduzir o viés de método. Não há, portanto, uma resposta única, e a escolha feita precisa ser justificada no texto.

A quarta é a sequência dos itens. Tourangeau e Rasinski (1988) relatam dois casos que mostram o alcance do efeito de ordem, isto é, da influência que a posição de uma pergunta exerce sobre a resposta a outra. No primeiro, uma pergunta sobre o direito ao aborto legal quando há forte probabilidade de malformação congênita grave, apresentada antes, reduziu as respostas favoráveis a outra pergunta, sobre o mesmo direito para uma mulher casada que não deseja mais filhos. No segundo, cerca de metade dos respondentes concordou que repórteres de países comunistas deveriam poder trabalhar nos Estados Unidos quando essa pergunta vinha primeiro; quando ela vinha depois de uma pergunta sobre repórteres americanos em países comunistas, a concordância subiu para quase três quartos. Os autores atribuem o segundo efeito à norma de tratar os dois lados da mesma forma, ativada pela pergunta anterior, e consideram efeitos como esses, em geral, produtos do processo de medição, e não da opinião que as perguntas pretendiam captar. Para o aluno, a consequência é testar a ordem durante o pré-teste e decidir conscientemente qual item vem primeiro quando dois deles tratam de objetos comparáveis. Podsakoff et al. (2003) sugerem preparar versões do questionário com os itens em ordens diferentes (contrabalançar), mas lembram que isso pode quebrar a progressão lógica do geral para o específico.

Na etapa 6, o instrumento concebido nas etapas anteriores é posto à prova. Beatty e Willis (2007) tratam de uma técnica para isso, a entrevista cognitiva, em que o pesquisador aplica o instrumento a poucas pessoas e investiga como cada uma interpretou os itens e chegou às respostas. Os autores descrevem duas formas de conduzi-la. Na primeira, o participante pensa em voz alta enquanto responde. Na segunda, o entrevistador faz perguntas complementares (sondagem), por exemplo, “o que este termo significa para você?”. Essas perguntas, como os próprios itens, não devem sugerir uma resposta “correta”. As duas formas costumam ser combinadas, e os autores observam que o peso de cada uma pode variar conforme o perfil dos participantes e o tema do questionário. A orientação usual é trabalhar em rodadas de 5 a 15 entrevistas, revisando os itens entre uma rodada e outra. Esse número é um ponto de partida, e não uma garantia, pois os autores relatam um estudo em que problemas sérios ainda apareciam

depois de 50 entrevistas. Para decidir quando parar, os autores não descrevem uma regra, e sim um julgamento. As entrevistas seguem até que as rodadas passem a revelar poucos problemas novos, embora uma entrevista a mais sempre possa trazer uma descoberta importante.

Abordagens que impactam a credibilidade do instrumento

Os erros que mais comprometem um instrumento raramente são de digitação ou de formatação. São decisões de concepção, de coleta e de análise tomadas sem prever o efeito que terão sobre os dados. O Quadro 2 reúne esses erros, identificados a partir dos nove artigos e organizados nessas três categorias com a consequência de cada um, a obra que fundamenta o diagnóstico e a etapa do Quadro 1 em que ele pode ser prevenido.

Quadro 2. Erros que comprometem a credibilidade do instrumento

Categoria	Erro	Consequência para os dados	Fonte	Etapa do Quadro 1
Concepção	Item que pede dois julgamentos ao mesmo tempo (dupla exigência).	O respondente decide a qual parte responder, e a resposta não pode ser interpretada.	Podsakoff et al. (2003)	3
Concepção	Termo técnico ou pouco familiar sem definição.	Cada respondente atribui ao item o próprio sentido, o que aumenta as respostas dadas ao acaso ou abre espaço para tendências pessoais.	Podsakoff et al. (2003); Tourangeau e Rasinski (1988)	3
Concepção	Item ou instrução que indica a resposta desejada.	A resposta passa a refletir o que o respondente supõe que o pesquisador espera.	Podsakoff et al. (2003)	3 e 4
Concepção	Análise de dados escolhida somente depois da coleta.	O tipo de dado gerado pela escala pode não permitir a análise pretendida.	Jamieson (2004)	1
Coleta	Ausência de pré-teste ou instrumento declarado 'validado' só porque passou por ele.	Problemas de compreensão só aparecem nos dados, quando já não podem ser corrigidos.	Beatty e Willis (2007)	6
Coleta	Comparação entre as autoavaliações de antes e de depois sem considerar que a intervenção pode mudar o critério com que o respondente se avalia.	A mudança medida pode não corresponder à mudança real. Os participantes de um curso podem, por exemplo, dar a si mesmos a mesma nota no início e no fim, não porque não aprenderam, mas porque o curso os fez perceber que sabiam menos do que pensavam no início.	Howard e Dailey (1979)	2
Coleta	Avaliação retrospectiva apresentada como livre de vieses.	Ao avaliar, no fim do curso, como se via antes dele, o respondente pode rebaixar essa nota para mostrar que o curso valeu a pena, o que faz a mudança parecer maior do que foi.	Howard e Dailey (1979)	2 e 7
Coleta	Itens parecidos colocados em sequência.	O respondente tende a manter coerência com as respostas anteriores, em vez de dizer o que pensa, sobretudo em relatos sobre o passado.	Tourangeau e Rasinski (1988); Podsakoff et al. (2003)	5
Coleta	Respondente que se sente avaliado.	O receio de ser avaliado leva a respostas socialmente mais aceitáveis, mais favoráveis ou de simples concordância.	Podsakoff et al. (2003)	7

Análise	Média e desvio-padrão em escala Likert sem explicar por que os pontos da escala foram tratados como se estivessem à mesma distância uns dos outros.	Distâncias que podem não ser iguais são tratadas como iguais, e, quando a maioria das respostas se concentra em um extremo da escala, poucas respostas do lado oposto puxam a média e a distorcem.	Jamieson (2004)	1, 4 e 8
Análise	Concluir que não há viés de método só porque o teste de fator único de Harman, que reúne todos os itens em uma análise fatorial, não encontrou um fator dominante.	O teste é pouco capaz de detectar o viés e não o corrige. Seu resultado não basta como prova, e a defesa contra o viés deve vir sobretudo dos cuidados de concepção e de coleta.	Podsakoff et al. (2003)	3 a 8
Análise	Temas que apenas repetem as perguntas do instrumento ou que se apoiam em um ou dois trechos de resposta.	No primeiro caso, as respostas são só reorganizadas, sem análise; no segundo, uma opinião isolada passa por padrão do grupo. Os trechos devem ilustrar padrões que aparecem em várias respostas.	Braun e Clarke (2006)	9
Análise	Apresentar a concordância entre uma escala e uma pergunta aberta do mesmo questionário como triangulação entre métodos (confirmação de um resultado por métodos diferentes).	Usar a resposta aberta para detalhar ou ilustrar a fechada é legítimo. Mas as duas vêm da mesma pessoa e do mesmo momento e podem conter o mesmo viés, como o desejo de agradar ao pesquisador, de modo que a concordância entre elas reforça o resultado, mas não o comprova.	Jick (1979)	2 e 10

Fonte: Elaboração própria, a partir dos autores citados (2026).

Dois desses erros merecem atenção especial, porque costumam passar despercebidos por quem elabora o instrumento.

O primeiro é a instrução que sugere a resposta. Podsakoff et al. (2003) recomendam identificar e eliminar dos itens as pistas sobre o que o pesquisador espera ouvir (características de demanda). Essas pistas podem estar no enunciado, mas também nas instruções e nos exemplos de preenchimento. Um exemplo que mostra notas baixas para o "antes" e notas altas para o "depois" informa ao respondente o padrão esperado antes que ele responda. O exemplo de preenchimento deve tratar de um assunto alheio à pesquisa e admitir qualquer resultado, inclusive a ausência de mudança e uma avaliação do "depois" abaixo da avaliação do "antes", como em "Antes do curso, eu sabia usar planilhas eletrônicas" e "Depois do curso, sei usar planilhas eletrônicas", ambos marcados com a nota 3.

O mesmo cuidado vale para as perguntas abertas. Perguntas do tipo "de que forma o protocolo ajudou" partem do princípio de que ele ajudou e deixam pouco espaço para quem pensa o contrário. No instrumento do workshop, a pergunta sobre a aplicabilidade do protocolo evitou esse problema ao oferecer as duas possibilidades:

> *O protocolo se mostra diretamente aplicável às particularidades da sua subárea específica em Administração (...), ou exigiria adaptações metodológicas importantes? Explique como o protocolo se ajusta à sua área ou, se for o caso, quais adaptações seriam requeridas.*

Formulada dessa maneira, a pergunta deixa o respondente tão à vontade para apontar limites quanto para confirmar a utilidade do protocolo.

Quando o próprio tema carrega forte pressão social, cuidar da redação não basta. Ao estudar a ansiedade de empregados durante uma fusão, Jick (1979) reconheceu que perguntar

diretamente sobre ela favorecia respostas ajustadas ao que os empregados imaginavam ser esperado. Por isso combinou questionários e entrevistas com medidas que não dependiam do que os empregados diziam, como a observação e o registro de suas visitas à biblioteca de arquivos da organização, onde eles buscavam informações sobre a fusão. O autor tratou essas visitas como um termômetro da ansiedade. Para o aluno, a lição é que, em temas sensíveis, vale complementar o questionário, sempre que possível, com dados que não dependam da resposta do participante.

O segundo erro é apresentar a avaliação retrospectiva como livre de vieses. Howard e Dailey (1979) mostraram que a mudança medida com ela concordou mais com a avaliação externa do que a medida com o pré-teste tradicional, mas reconhecem que não descartaram outras explicações. Um participante satisfeito com o curso, por exemplo, pode rebaixar a avaliação do "antes" para mostrar ao pesquisador que o curso valeu a pena, tendência que os autores associam à aquiescência. No estudo deles, as comparações entre avaliação retrospectiva e pós-teste mostraram 126 casos de melhora, 21 sem mudança e nenhum de piora, um resultado tão uniforme que pede cautela na interpretação.

Podsakoff et al. (2003) observam que a busca por coerência entre as respostas é particularmente problemática quando se pedem relatos sobre o passado, e que o receio de ser avaliado leva o respondente a ajustar a resposta na direção do que considera desejável. Esse risco tende a crescer quando quem aplica o instrumento também conduziu a intervenção ou avalia o desempenho do respondente. Os remédios que os autores propõem incluem garantir o anonimato, informar que não há respostas certas ou erradas e, quando possível, coletar as medidas em momentos, locais ou meios diferentes, ou com respondentes diferentes.

Os próprios autores reconhecem que separar as medidas tem custos. A separação exige mais tempo, esforço e recursos, pode permitir que outros fatores interfiram entre uma medida e outra e, se o intervalo for longo, aumenta a perda de respondentes. Para quem não pode separar a coleta, situação comum em dissertações e teses, Podsakoff et al. (2003) indicam uma sequência pragmática. Os cuidados com a redação dos itens devem ser adotados sempre, e os autores afirmam não ver desvantagem neles. Em seguida vêm o anonimato e a garantia de que não há respostas certas, já mencionados, e a separação psicológica, que consiste em apresentar uma medida como se não tivesse relação com a outra. Quando a pesquisa testa relações entre construtos e esses cuidados não afastam o risco, os autores recomendam controles estatísticos e advertem que o teste de fator único de Harman, muito usado, não controla o viés nem demonstra sua ausência. O objetivo, afirmam, não é eliminar todo viés de método, o que pode ser impossível, mas reduzir a plausibilidade de que ele explique os resultados.

Da coleta à análise

A forma de cada item determina o que se pode afirmar com ele. Jamieson (2004) lembra que as categorias de uma escala Likert têm ordem, mas não se pode presumir que as distâncias entre elas sejam iguais, o que faz desses dados ordinais. A autora parte de um problema concreto, observado em dois artigos publicados no mesmo periódico em que o trabalho foi divulgado, que usaram escalas Likert, descreveram os dados por médias e desvios-padrão, e aplicaram análise de variância. Ela observa que, em geral, os autores não declaram que tratam esses dados como se as distâncias fossem iguais nem justificam essa escolha. Para dados ordinais, recomenda descrever os resultados pela mediana ou pela moda e pelas frequências de cada categoria, e usar testes não paramétricos, como o qui-quadrado e o teste U de Mann-Whitney. Acrescenta que, mesmo para quem aceita tratar as distâncias como iguais, esses dados

costumam ter distribuição assimétrica ou polarizada, como quando a maioria dos estudantes concorda que uma intervenção foi útil. Ao final do texto, a autora se declara convencida por um argumento que resume assim:

> *a média de "regular" e "bom" não é "regular e meio"; isso continua verdadeiro mesmo quando se atribuem números inteiros para representar "regular" e "bom". (Jamieson, 2004, p. 1218,* tradução nossa*)*

Jamieson (2004) reconhece que o tema continua controverso, mas defende que a decisão sobre o nível de mensuração, isto é, tratar a escala como ordinal ou como se as distâncias entre as categorias fossem iguais, seja tomada ainda no desenho da pesquisa e discutida pelos autores ao apresentar o método. Por isso essa decisão aparece na etapa 1 do roteiro (Quadro 1). O mesmo raciocínio vale para o cálculo da mudança de cada respondente. Subtrair a nota do "antes" da nota do "depois" é tratar as distâncias entre as categorias como iguais, e a dissertação ou a tese precisa declarar e justificar essa escolha.

Conover e Iman (1981) oferecem uma ponte entre a análise paramétrica, baseada em médias, e a não paramétrica. Os autores mostram que vários testes não paramétricos, entre eles o teste de Wilcoxon para dados pareados, podem ser obtidos aplicando o procedimento paramétrico usual aos postos dos dados, isto é, às posições que os valores ocupam quando ordenados do menor para o maior, em vez dos valores originais. No caso pareado, as diferenças entre as duas medidas de cada respondente são ordenadas pelo valor absoluto e recebem o sinal da diferença. Os autores registram que as tabelas exatas só podem ser usadas quando a amostra é pequena e não há empates. Havendo empates, os valores empatados recebem a média dos postos que ocupariam, e a estatística apresentada pelos autores já incorpora a correção correspondente. Em escalas de cinco pontos, empates e diferenças nulas são frequentes, e o aluno deve declarar como os tratou.

O tamanho da amostra impõe outro limite, que o aluno precisa conhecer antes da coleta. Imagine, por exemplo, um mestrando que aplica o instrumento a uma turma de oito participantes, antes e depois de um curso, e verifica que a maioria melhorou a avaliação. Para afirmar que essa melhora não foi obra do acaso, ele recorre a um teste estatístico, que devolve um valor de p. Esse valor indica a probabilidade de encontrar, apenas por acaso, uma diferença igual ou maior que a observada, caso o curso não tivesse efeito algum. Quanto menor o valor, menos plausível é atribuir a melhora ao acaso, e costuma-se exigir que ele fique abaixo de 0,05, ou 5%. O que muitos alunos não sabem é que, em amostras pequenas, o teste de Wilcoxon tem um valor de p mínimo, abaixo do qual não consegue chegar, por maior que seja a melhora.

Em um cálculo que fizemos para ilustrar este ponto, com os oito participantes e teste bilateral, que considera mudança em qualquer direção, esse mínimo é 0,0078 (0,78%) quando os oito mudam de avaliação no mesmo sentido. Os que mantêm a avaliação geram diferença nula e, neste cálculo, foram excluídos, o que reduz a amostra efetiva. Se dois mantêm a avaliação e seis mudam, o mínimo sobe para 0,031 (3,1%), ainda abaixo de 5%. Se três mantêm e cinco mudam, sobe para 0,0625 (6,25%), acima de 5%. Nesse caso, mesmo que os cinco tenham melhorado muito, o teste não indicará diferença significativa, e o mestrando não poderá afirmar que o curso funcionou, embora a melhora possa ser real. Por isso, em amostras desse tamanho, o valor de p deve vir acompanhado da contagem de quantas avaliações melhoraram, se mantiveram ou pioraram, como fizeram Howard e Dailey (1979) ao comparar cada item de cada participante.

Também nas perguntas abertas a análise depende de decisões que precisam ser tomadas e declaradas. Braun e Clarke (2006) descrevem a análise temática em seis fases: familiarização com os dados, geração de códigos iniciais (rótulos atribuídos a trechos das respostas), busca de temas, revisão dos temas, definição e nomeação dos temas e produção do relatório. As autoras pedem que o pesquisador explicite suas escolhas, por exemplo se os temas partem dos próprios dados ou de uma teoria e de perguntas de pesquisa definidas antes, e se a análise fica no que os respondentes disseram explicitamente ou busca as ideias e os pressupostos que estão por trás do que disseram. Entre as armadilhas que apontam, duas são comuns em instrumentos com perguntas abertas. A primeira é usar as próprias perguntas como temas. Se o questionário pergunta o que o participante mais valorizou no curso e o relatório apresenta o tema "Aspectos mais valorizados", as respostas foram apenas agrupadas, sem análise que identifique padrões no conjunto dos dados. A segunda é apresentar um ou dois trechos de resposta por tema, o que não basta para convencer quem não leu todas as respostas e pode fazer casos isolados parecerem padrão. As autoras criticam também a afirmação de que os temas "emergiram" dos dados, porque ela apaga o papel ativo do pesquisador, que é quem identifica, seleciona e apresenta os temas. Esse papel aparece nas perguntas que elas sugerem para as fases finais da análise, como o que cada tema significa, que pressupostos o sustentam e que história o conjunto dos temas conta sobre o assunto estudado. Para conferir o resultado, oferecem uma lista de 15 critérios de uma boa análise temática.

Resta integrar os dois tipos de dado. Jick (1979) distingue a triangulação dentro de um mesmo método, como o uso de várias escalas em um questionário, que verifica sobretudo a consistência interna, da triangulação entre métodos diferentes, que busca validação convergente. Para o autor, a limitação da primeira está justamente em usar um único método. Por esse critério, uma escala e uma pergunta aberta respondidas pela mesma pessoa, no mesmo questionário e no mesmo momento, pertencem ao mesmo método e compartilham seus vieses. Isso não reduz o valor das respostas abertas, mas limita o que se pode concluir quando elas confirmam os números.

No mesmo estudo, Jick (1979) mostra o valor da divergência. O autor esperava que os mais ansiosos frequentassem mais a biblioteca, mas os respondentes que relatavam mais ansiedade no questionário eram os que menos a visitavam. Entrevistas e observações adicionais indicaram que esses empregados, com menor escolaridade, se informavam mais pela comunicação oral do que por documentos escritos. Para o autor, a divergência pode ser uma oportunidade de enriquecer a explicação, e foi o que aconteceu nesse caso. Jick classifica ainda como forma primitiva de triangulação o uso parentético, e até condescendente, de observações de campo para reforçar resultados estatísticos. No exemplo hipotético que apresenta, um estudo sobre satisfação no trabalho mostra um teste qui-quadrado que indica forte descontentamento e, para reforçá-lo, menciona que houve uma greve naquele ano, sem informar a intensidade, a dinâmica, o significado e as consequências da greve. Nesse caso, o dado qualitativo não foi integrado ao quantitativo; serviu apenas de ilustração.

Quando o objetivo é explicar os resultados quantitativos, um desenho adequado é o sequencial explicativo descrito por Ivankova et al. (2006). Nele, o pesquisador coleta e analisa primeiro os dados quantitativos e só depois realiza a fase qualitativa, cujos participantes e cujo roteiro são definidos a partir dos resultados da primeira fase. No estudo dos autores, sobre a permanência de estudantes em um doutorado a distância, a fase qualitativa examinou um caso típico de cada um dos quatro grupos de estudantes, escolhido a partir das pontuações obtidas na fase quantitativa, e o roteiro das entrevistas foi construído a partir dos fatores que a análise

estatística havia destacado. Os autores apontam como limitações desse desenho o tempo e os recursos necessários para coletar e analisar os dois tipos de dado. Perguntas abertas incluídas no mesmo questionário das escalas não configuram esse desenho. Se o aluno quiser adotá-lo, precisa planejar a segunda fase desde a concepção do instrumento e verificar se o prazo do curso comporta as duas fases.

As escolhas feitas no desenho do instrumento determinam o que o aluno poderá afirmar ao final da pesquisa. Voltemos ao exemplo do curso. Se o instrumento pediu que os participantes se autoavaliassem, o resultado mostra como eles percebem a própria competência, e não se ela de fato aumentou. Para afirmar que o curso melhorou a competência dos participantes, é preciso outro tipo de evidência, como a avaliação feita por juízes no estudo de Howard e Dailey (1979). Da mesma forma, os resultados obtidos com uma turma pequena descrevem aquela turma e não podem ser estendidos a todos os participantes de cursos semelhantes. Concluir além do que os dados permitem é uma das críticas mais previsíveis em bancas. Para evitá-la, o aluno deve decidir, já na etapa 1 do roteiro, que tipo de afirmação pretende fazer e construir um instrumento capaz de sustentá-la.

Considerações finais sobre a elaboração de instrumentos

Ao discutir os limites da entrevista cognitiva, Beatty e Willis (2007) fazem uma advertência que vale para o instrumento como um todo:

> afirmações de que um questionário "não tem problemas" são impossíveis; o máximo que se pode afirmar é que nenhum problema foi (ainda) descoberto. (Beatty & Willis, 2007, p. 303, tradução nossa)

O propósito deste editorial não é oferecer uma fórmula que torne o instrumento imune a críticas, e sim um roteiro que permita ao aluno mostrar que cada decisão foi tomada com conhecimento de suas consequências.

Algumas falhas aparecem com mais frequência nas bancas de qualificação de que participamos. A mais comum, já comentada, é escolher a análise depois da coleta. Logo depois vêm os itens que pedem dois julgamentos ao mesmo tempo, como "O curso foi útil e bem organizado?", ou que indicam ao respondente a resposta esperada. Por fim, há a apresentação do instrumento como validado, ou da triangulação como prova, sem as ressalvas feitas pelos próprios autores em que o trabalho se apoia.

Para evitar essas e as demais falhas discutidas ao longo deste editorial, reunimos suas recomendações em um checklist para quem vai elaborar e aplicar um questionário (Quadro 3). As perguntas, de sim ou não, permitem ao aluno verificar o próprio instrumento em dois momentos, antes de levá-lo a campo e, depois, ao analisar e relatar os resultados. Os três quadros funcionam em conjunto. O Quadro 1 orienta cada decisão, o Quadro 2 mostra o que acontece quando ela é negligenciada e o Quadro 3 permite conferir se ela foi tomada. Por isso, cada pergunta indica a etapa do Quadro 1 a que se refere, e quem responder não a uma delas sabe onde encontrar a orientação correspondente. Uma resposta negativa não impede a pesquisa, mas indica um ponto a corrigir ou, quando não houver correção possível, a declarar como limitação.

Quadro 3. Checklist para o aluno aplicar ao próprio instrumento

Momento	Etapa	Pergunta de verificação	Sim	Não
Antes de redigir	1	Cada item ajuda a responder ao problema de pesquisa?		
Antes de redigir	1	Sei que tipo de dado cada item vai gerar e qual análise farei com ele?		
Antes de redigir	2	Defini quando e quantas vezes vou aplicar o instrumento e se as perguntas fechadas e abertas serão aplicadas juntas ou em fases separadas?		
Antes de redigir	2	Se vou medir mudança pela autoavaliação do respondente, considerei que o próprio curso ou intervenção pode mudar a forma como ele se avalia?		
Antes de redigir	2 e 8	O número de respondentes permite que o teste escolhido alcance um valor de p abaixo de 0,05?		
Antes de redigir	2 e 10	Se pretendo usar entrevistas ou perguntas abertas para explicar os resultados das questões fechadas, elas serão aplicadas depois da análise desses resultados e formuladas a partir deles?		
Redação	3	Cada item pede uma única avaliação, sem juntar dois assuntos na mesma pergunta?		
Redação	3	Os termos que o respondente pode não conhecer estão explicados no próprio item?		
Redação	3 e 4	Os itens, as instruções e os exemplos de preenchimento evitam sugerir a resposta que eu espero?		
Escala	4	Todos os pontos da escala têm nome, inclusive o do meio (por exemplo, “nem concordo nem discordo”)?		
Escala	4	Evitei escalas numéricas bipolares, do tipo −3 a +3, a menos que o construto seja bipolar por natureza?		
Sequência	5	Verifiquei se a resposta a um item pode influenciar a resposta aos seguintes?		
Pré-teste	6	Antes da coleta, verifiquei com algumas pessoas do público pesquisado como elas entendem cada item (entrevista cognitiva) e registrei as alterações feitas?		
Aplicação	7	O respondente será informado de que a pesquisa é anônima e de que não há respostas certas ou erradas?		
Aplicação	7	A forma de aplicação evita que o respondente se sinta avaliado ou queira agradar o pesquisador, como quando quem aplica é alguém próximo?		
Aplicação	7	Se vou comparar respostas do mesmo respondente dadas em momentos diferentes, previ um código que as ligue sem identificá-lo?		
Análise	8	Justifiquei o uso da mediana e de testes não paramétricos, ou da média, e informei como tratei os empates (valores repetidos) e os respondentes que não mudaram de avaliação?		
Análise	9	Os temas das respostas abertas resultam da análise das respostas, e não apenas repetem as perguntas que fiz?		
Análise	9	Cada tema identificado nas respostas abertas aparece em várias respostas, e não apenas em uma ou duas?		
Análise	9	Nas fases finais da análise, perguntei o que cada tema significa, que pressupostos o sustentam e que história o conjunto dos temas conta?		
Integração	10	Expliquei os pontos em que as respostas abertas divergem dos resultados numéricos, e não apenas aqueles em que concordam?		
Integração	10	Tratei a concordância entre escala e pergunta aberta do mesmo questionário como reforço do resultado, e não como prova dele?		
Relato	1 a 10	Guardei as versões sucessivas do instrumento e a justificativa de cada mudança?		
Relato	1 a 10	Descrevi os limites do instrumento, como fazem os próprios autores em que me apoiei?		
Relato	1	As conclusões se limitam ao que o instrumento permite afirmar (por exemplo, à percepção dos participantes, quando se usa autoavaliação)?		

Fonte: Elaboração própria (2026).

Recomendamos também que o aluno guarde as versões sucessivas do instrumento, o registro das entrevistas cognitivas e a justificativa de cada mudança, que alimentam a última coluna do Quadro 1. Esse material raramente cabe no corpo de uma dissertação ou de um artigo, mas pode compor um apêndice. Ele responde, com antecedência, às perguntas que bancas ou pareceristas costumam fazer sobre como o instrumento foi construído.

Tudo isso cabe no prazo de um mestrado? É provável que o leitor já tenha se feito essa pergunta, e a resposta depende de quando o instrumento entra no projeto. Boa parte do que foi discutido aqui custa pouco tempo se for decidida na fase de desenho. Definir o construto, escolher a análise, verificar se o número de respondentes comporta o teste pretendido, redigir itens simples e sem dupla exigência e rotular a escala são decisões tomadas antes da coleta. Para os cuidados de redação, Podsakoff et al. (2003) afirmam não ver desvantagem alguma. O que exige tempo são as etapas que dependem de campo, como as rodadas de entrevista cognitiva, a separação das medidas e a segunda fase de um desenho sequencial. São essas que o aluno precisa dimensionar diante do calendário. Braun e Clarke (2006) incluem, entre os critérios de uma boa análise temática, o de reservar tempo suficiente para todas as fases da análise, sem atropelar nenhuma. Quando não for possível cumprir uma etapa, o aluno não deve omiti-la, e sim registrar a escolha e o que ela limita. Como na advertência de Beatty e Willis (2007) que abriu estas considerações, nenhum instrumento pode ser declarado sem problemas, e o que se espera de uma dissertação (ou tese) é que o aluno saiba quais são os seus.

Apresentação dos artigos desta edição

Apresentamos a seguir os seis trabalhos deste número, na ordem em que aparecem. O primeiro artigo, “Alcance de metas do Plano Nacional de Resíduos Sólidos de 2022: caso de dois municípios da região Sul Fluminense do Estado do Rio de Janeiro”, mostra, com dados de 2015 a 2021 do Sistema Nacional de Informações sobre Saneamento, que Barra Mansa e Volta Redonda superaram as metas de coleta domiciliar, mas só Barra Mansa avançou na coleta seletiva e na recuperação de recicláveis.

A mobilidade urbana sustentável é o tema de “Avaliação de diretrizes de impacto de tráfego aplicadas a PGVs: um estudo internacional com ênfase no uso da bicicleta”. O estudo comparou cinco diretrizes internacionais quanto à integração da bicicleta ao planejamento dos polos geradores de viagens e apontou Singapura como a mais alinhada à promoção do ciclismo. Segundo os autores, o modelo de análise pode ser replicado no Brasil.

O terceiro trabalho é o caso de ensino “Santo de casa [não] faz milagre? Os desafios de Olavo na reestruturação de uma franquia”. O caso acompanha um profissional prestes a se aposentar que assume o RH de uma franquia do ramo de alimentação, e seu dilema, entre se aposentar e assumir mais responsabilidades, abre a discussão sobre as dores do crescimento de uma empresa.

Em “Atividades de inovação orientada à sustentabilidade: um estudo no contexto de empresas que atuam em *clusters* industriais brasileiros”, os autores mostram que a proximidade geográfica favorece inovações voltadas à otimização das operações, enquanto as redes de colaboração e as instituições de apoio viabilizam parcerias com universidades e outros agentes.

Com base em 104 artigos da *Web of Science* publicados entre 2015 e 2024, “O uso de tecnologias na formação de estratégias competitivas e de inovação em serviços: uma revisão bibliométrica” identifica quatro núcleos temáticos, a servitização digital, a experiência do cliente, as capacidades dinâmicas e a inovação aberta.

O número termina com "Fundamentos de Big Data no sistema de justiça: o uso de dados na Defensoria Pública do Estado do Amazonas", um estudo de caso, com pesquisa documental e entrevistas, sobre o uso estruturado de dados pelo órgão desde 2018. Com o apoio da alta gestão, essa prática melhorou o desempenho do órgão e fez com que as decisões, antes baseadas na experiência, passassem a se apoiar em dados.

Agradecemos aos autores, aos avaliadores, aos editores associados e a toda a equipe editorial pelo trabalho que tornou este número possível. Desejamos aos nossos leitores boas festas e um próspero ano novo. Até o próximo editorial, que em 2027 abrirá o décimo terceiro volume da Revista de Administração, Sociedade e Inovação.

Declaração de uso de inteligência artificial

Os autores utilizaram o Claude (Anthropic) como ferramenta de apoio na redação e na revisão deste editorial. A concepção, a seleção das fontes, as decisões de conteúdo e a revisão final do texto são dos autores, que assumem integral responsabilidade pelo conteúdo publicado.

Referências

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 71(2), 287–311. https://doi.org/10.1093/poq/nfm006

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35(3), 124–129. https://doi.org/10.1080/00031305.1981.10479327

Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, 64(2), 144–150. https://doi.org/10.1037/0021-9010.64.2.144

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, 18(1), 3–20. https://doi.org/10.1177/1525822X05282260

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, 38(12), 1217–1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 24(4), 602–611. https://doi.org/10.2307/2392366

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103(3), 299–314. https://doi.org/10.1037/0033-2909.103.3.299