

TEXTOS PARA DISCUSSÃO

DIRETORIA DE PESQUISAS

n.57

CONTROLE DE QUALIDADE

EM PESQUISAS

OS TESTES COMO INSTRUMENTO
DE AVALIAÇÃO E APRIMORAMENTO
DE QUESTIONÁRIOS

Andréa Borges Paim

Presidente da República
Jair Messias Bolsonaro

Ministro da Economia
Paulo Roberto Nunes Guedes

Secretário Especial de Fazenda
Waldery Rodrigues Junior

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE

Presidente
Susana Cordeiro Guerra

Diretor-Executivo
Fernando José de Araújo Abrantes

ÓRGÃOS ESPECÍFICOS SINGULARES

Diretoria de Pesquisas
Eduardo Luiz G. Rios Neto

Diretoria de Geociências
João Bosco de Azevedo

Diretoria de Informática
David Wu Tai

Centro de Documentação e Disseminação de Informações
Marise Maria Ferreira

Escola Nacional de Ciências Estatísticas
Maysa Sacramento de Magalhães

UNIDADE RESPONSÁVEL

Diretoria de Pesquisas
Coordenação de Métodos e Qualidade
Marcus Vinicius Morais Fernandes (em exercício)

Ministério da Economia
Instituto Brasileiro de Geografia e Estatística – IBGE
Diretoria de Pesquisas
Coordenação de Métodos e Qualidade

Textos para discussão
Diretoria de Pesquisas
número 57

Controle de qualidade em pesquisas: os testes como instrumento de avaliação e aprimoramento de questionários

Andréa Borges Paim

Instituto Brasileiro de Geografia e Estatística - IBGE
Av. Franklin Roosevelt, 166 - Centro - 20021-120 - Rio de Janeiro, RJ – Brasil

Textos para discussão. Diretoria de Pesquisas, ISSN 1518-675X
Divulga estudos e outros trabalhos técnicos desenvolvidos pelo IBGE ou em conjunto com outras instituições, bem como resultantes de consultorias técnicas e traduções consideradas relevantes para disseminação pelo Instituto. A série está subdividida por unidade organizacional e os textos são de responsabilidade de cada área específica.

ISBN 978-85-240-4497-7

© IBGE. 2019

Impressão

Gráfica Digital/Centro de Documentação e Disseminação de Informações - CDDI/IBGE, em 2019.

Capa

Gerência de Criação/CDDI

Ficha catalográfica elaborada pela Gerência de Biblioteca e Acervos Especiais do IBGE

Paim, Andréa Borges

Controle de qualidade em pesquisas : os testes como instrumento de avaliação e aprimoramento de questionários / Andréa Borges Paim. - Rio de Janeiro : IBGE, 2019.

32 p. - (Textos para discussão. Diretoria de Pesquisas, ISSN 1518-675X, n. 57)

Inclui bibliografia e glossário.

ISBN 978-85-240-4497-7

1. Questionários. 2. Métodos estatísticos. 3. Metodologia I. IBGE. Coordenação de Métodos e Qualidade. II. Título. III. Série.

CDU 311.211

EST

Sumário

APRESENTAÇÃO	5
INTRODUÇÃO	7
ETAPAS LÓGICAS DA CONSTRUÇÃO DE UM QUESTIONÁRIO	8
TESTES COMO INSTRUMENTO DE AVALIAÇÃO E APRIMORAMENTO DE INSTRUMENTOS DE COLETA.....	11
Métodos para testes pré-campo.....	12
Revisões Metodológicas	14
Grupos Focais.....	15
Entrevistas Cognitivas.....	16
Testes de Usabilidade	18
Experimentos Aleatorizados	19
Confiabilidade e Validade de um instrumento de coleta	20
Recomendações gerais para realização de testes de questionário	21
Pesquisas novas.....	22
Pesquisas em andamento.....	25
Questionário do próprio	26
TESTES DE QUESTIONÁRIO REALIZADOS NO IBGE	27
CONSIDERAÇÕES FINAIS	28
REFERÊNCIAS.....	30

Apresentação

Os dados obtidos pelos Institutos Oficiais de Estatística são importantes para subsidiar as decisões referentes às políticas públicas e privadas em prol da sociedade. E, para tal, faz-se necessário assegurar a qualidade do instrumento de coleta, que deve ser elaborado de forma a atender o objetivo da pesquisa, captando os fenômenos de interesse com a maior precisão possível.

Segundo as boas práticas internacionais em qualidade de processos estatísticos, um instrumento de coleta deve passar por etapas de teste e validação antes que se inicie a operação de coleta de dados.

Neste texto, são apresentados diversos métodos de testes que podem ser utilizados para avaliar se as perguntas elaboradas na etapa de construção do instrumento de coleta estão captando com qualidade a informação desejada.

Marcos Paulo Soares de Freitas
Coordenação de Métodos e Qualidade¹

¹ No preparo dos originais do presente volume, a **Apresentação** foi elaborada pelo então titular da Coordenação de Métodos e Qualidade, cujo ocupante atual, em exercício, é Marcus Vinicius Moraes Fernandes.

Introdução

Pode-se definir controle de qualidade em pesquisas como o conjunto de práticas e procedimentos implementados pelos institutos de pesquisa visando garantir não somente a precisão dos dados coletados como também que os mesmos estejam adequados para utilização e em conformidade com o propósito a que se destinam. Uma vez que os resultados das pesquisas de um Instituto Oficial de Estatística fornecem subsídios para tomada de decisões tanto na esfera pública quanto na privada, o controle de qualidade do processo de produção estatística deve ser implementado de forma sistemática para minimizar erros e prevenir práticas que afetem os resultados das pesquisas, abrangendo dimensões como relevância, acurácia, pontualidade, coerência, comparabilidade e acessibilidade. A escolha da melhor metodologia de controle de qualidade varia conforme o tipo de pesquisa e sua aplicação se destina a todos os aspectos do processo de pesquisa, como questionários, amostragem, gerenciamento da coleta e monitoramento do comportamento adequado do entrevistador.

De acordo com o DataONE (2016), Quality Assurance (Garantia de Qualidade) e Data Quality Control (Controle de Qualidade de Dados) são estratégias adotadas para evitar erros em um conjunto de dados, sendo que a primeira está relacionada com as atividades para assegurar a qualidade dos dados antes da coleta e a segunda, ao monitoramento e à manutenção da qualidade dos dados durante a condução da pesquisa. Já Üstun, Chatterji, Mechbal e Murray (2005) entendem o conceito de Quality Assurance como “uma ferramenta organizacional para implementação com padrões operacionais pré-definidos em relação à estrutura, processo e resultados da pesquisa”, com utilidade para coleta de dados de alta qualidade e validação dos resultados. Dessa forma, Quality Assurance pode ser vista como um processo contínuo que ocorre em todas as fases da pesquisa, desde a preparação e amostragem, passando pela coleta e análise de dados, até a elaboração dos relatórios finais.

Um dos pilares de uma pesquisa bem-sucedida é o instrumento de coleta, que deve ser elaborado em um fluxo lógico de forma a atender aos objetivos da pesquisa e minimizar a taxa de perguntas não respondidas. Dessa forma, as conclusões da pesquisa sobre o fenômeno observado serão precisas e confiáveis, havendo, portanto, garantia de qualidade. Visando auxiliar na elaboração dos instrumentos de coleta, o presente texto aborda a utilização dos testes como uma forma de avaliação e aprimoramento. Cabe mencionar que o estudo em questão está em conformidade com o Princípio 10 do Código de Boas Práticas das Estatísticas do IBGE, sobre processos estatísticos adequados, em que “o IBGE deve utilizar, em todas as etapas do processo estatístico, procedimentos e ferramentas adequados para garantir a qualidade das estatísticas oficiais”. Em relação a esse princípio, destaca-se ainda, como indicador de boas práticas, que “as metodologias, os questionários, os manuais, os sistemas de informática e os demais instrumentos devem ser testados e validados antes do início do processo de coleta de dados” (IBGE, 2013).

Etapas lógicas da construção de um questionário

Segundo Marconi e Lakatos (1999), o questionário é um instrumento desenvolvido cientificamente que visa coletar informações de forma sistemática e ordenada sobre temas de interesse em uma população de estudo. Em geral, os questionários são utilizados na obtenção de grandes quantidades de dados e se constituem na espinha dorsal da pesquisa. Portanto, o processo de construção de um questionário demanda tempo, disponibilidade e interesse por parte dos pesquisadores, em especial por não existir uma metodologia padrão que garanta que seu objetivo seja alcançado com qualidade. Não obstante, para se construir um questionário eficaz, é essencial seguir uma sequência de etapas lógicas, conforme ilustra o Quadro 1 a seguir:

Quadro 1. Sequência de etapas lógicas e passos a serem seguidos na elaboração de um questionário.

Etapa Lógica	Passos
Planejar o que vai ser mensurado	Evidenciar os objetivos da pesquisa
	Definir o assunto da pesquisa e as hipóteses em torno das quais se deve projetar o questionário
	Obter informações adicionais sobre o tema da pesquisa a partir de fontes de dados secundários e pesquisa exploratória
	Determinar o que será perguntado sobre o tema
	Definir a população-alvo
Dar forma ao questionário	Para cada tema, determinar o total de perguntas e o conteúdo de cada pergunta
	Decidir sobre o formato de cada pergunta
Elaborar textos das perguntas	Determinar como as questões serão redigidas ou buscar perguntas já validadas em bancos de perguntas ou em questionários validados
	Avaliar cada uma das questões em termos de facilidade de compreensão, disposição, conhecimentos e habilidades exigidos dos respondentes
Decidir sequenciamento e aparência	Dispor os temas e as questões em uma ordem adequada
	Agrupar todas as questões de cada subtópico para obter um único questionário
Pré-teste e correção de problemas	Ler o questionário inteiro para verificar se faz sentido e se o questionário efetivamente consegue mensurar o fenômeno a que se propõe
	Verificar possíveis erros no questionário (grafia, sequência, entre outros)
	Fazer os testes do questionário
	Corrigir os problemas detectados

Fonte: Adaptado de Aaker et al. (2001).

Em relação à elaboração das perguntas do questionário, deve-se ter uma avaliação

crítica da real necessidade e contribuição de cada pergunta para atingir o objetivo da pesquisa. Se a pergunta em questão não contribui diretamente para testar uma ou mais hipóteses relacionadas ao tema de estudo, deve ser retirada do questionário. São exceções aquelas perguntas que servem como uma “abertura” do tema, que sejam fáceis de responder e cujo objetivo seja estabelecer uma ligação com o entrevistador ou envolver o respondente na pesquisa.

Pode-se verificar a necessidade de manter uma determinada pergunta em um questionário através dos seguintes questionamentos:

1. A pergunta entra em conflito ou duplicidade com alguma outra já existente no questionário?
2. O texto da pergunta é específico ou minucioso em excesso, de forma a dificultar a elaboração da resposta?
3. As referências utilizadas na pergunta são claras e permitem um entendimento uniforme por parte dos respondentes?
4. A elaboração da pergunta atende aos aspectos importantes sobre o tema?
5. O respondente possui a informação necessária para responder à pergunta?
6. A pergunta pode trazer algum tipo de desconforto ao respondente, causando respostas que não reflitam a verdade ou mesmo recusa? (Em geral, ocorre com temas ligados à intimidade das pessoas, conteúdos emocionais, assuntos perturbadores, pontos de vista contrários às ideias socialmente aceitas ou ainda que gerem exposição social ou econômica)
7. O texto da pergunta foi elaborado de forma neutra, eliminando viés ou direcionamento da resposta em determinada direção?

Da mesma forma, deve-se verificar se as perguntas têm o mesmo significado para o entrevistador e para o respondente, evitando um erro de medição. Para tal, a linguagem utilizada deve ser simples, sem termos rebuscados ou obscuros e sim termos bem compreendidos pela população-alvo, assim como sentenças simples e curtas. Deve-se ainda evitar a junção de duas perguntas em uma só, causando imprecisão na resposta. Complementos da pergunta após seu texto principal tendem a confundir o respondente e também devem ser evitados.

Outro ponto importante é a sequência das perguntas, que pode influenciar na qualidade das respostas. Um primeiro contato bem-sucedido do respondente com o questionário faz com que o primeiro se sinta mais receptivo e disposto a colaborar com a pesquisa. Dessa forma, recomenda-se iniciar o questionário com perguntas gerais e conduzir gradualmente o foco para o tema central da pesquisa. Perguntas pessoais, sensíveis ou que causem constrangimento devem estar no final do questionário, preferencialmente alternadas com questões mais simples, para minimizar seu impacto.

Embora haja uma recomendação para posicionar no final do questionário as perguntas de classificação demográfica ou socioeconômica, em alguns casos essas perguntas servem como uma espécie de filtro, influenciando no fluxo do questionário. Nesse caso, essas

perguntas devem estar posicionadas de forma a permitir os “saltos” necessários à sequência do questionário. Ainda em relação ao fluxo do questionário, deve-se atentar à sequência lógica do questionário, evitando tanto mudanças bruscas de tema quanto retornos a temas abordados anteriormente.

A qualidade dos dados coletados também se relaciona com as características físicas do questionário. Recomenda-se que o mesmo não seja extenso, para não tornar seu preenchimento ou a entrevista cansativa para o respondente e evitar interrupções. O layout e a apresentação devem facilitar a leitura e compreensão das perguntas tanto por parte do entrevistador quanto do respondente, com formatos claros e simplificados, utilização criativa do espaço e das cores. Recomenda-se ainda que as instruções de preenchimento estejam próximas das perguntas.

O formato das respostas e sua redação também exercem influência sobre a qualidade dos dados, devendo obedecer aos mesmos critérios de elaboração e apresentação das perguntas. Em relação aos tipos de resposta, podem ser abertas ou fechadas e sua escolha deve considerar as vantagens e desvantagens conforme o objetivo da pesquisa.

Outro ponto importante é o tempo ideal da entrevista ou autopreenchimento do questionário. Estudos recentes mostram que o tempo médio de atenção de um adulto é estimado em cerca de 20 minutos (CAPE E PHILLIPS, 2015), indicando esse tempo como um tempo ideal de duração de entrevistas. Revilla e Ochoa (2017) conduziram um estudo com 755 respondentes, visando avaliar em quanto tempo (em minutos) levariam para responder um questionário de 115 perguntas e qual seria a duração ideal de uma entrevista. Os resultados do estudo ratificaram o tempo médio de 20 minutos como ideal para a duração do preenchimento de um questionário, uma vez que, após esse tempo, haveria perda de interesse dos respondentes em prestar as informações.

Testes como instrumento de avaliação e aprimoramento de instrumentos de coleta

Mesmo que as orientações apresentadas no Capítulo 2 tenham sido rigorosamente seguidas, não se pode afirmar que o questionário elaborado é um instrumento de coleta adequado para captar os dados desejados com a qualidade necessária até que o mesmo tenha sido efetivamente aplicado à população que se deseja estudar. Portanto, mediante a impossibilidade de prever de forma exaustiva os problemas que possam surgir durante a aplicação do questionário, faz-se necessário realizar testes para verificar se o mesmo irá atingir os objetivos da pesquisa com garantia de qualidade.

O termo “teste” abrange uma vasta gama de diferentes métodos que podem ser utilizados na avaliação da qualidade tanto da operacionalização de uma pesquisa quanto dos instrumentos de coleta. Embora com suas limitações, cada método tem sua utilidade na identificação de possíveis problemas nas diferentes fases do desenvolvimento de um instrumento de coleta. Portanto, conforme a complexidade da pesquisa, indica-se a combinação de um ou mais métodos para uma identificação e solução eficaz de problemas.

Somente através de testes é possível:

1. Identificar e corrigir possíveis erros;
2. Determinar se as perguntas foram elaboradas de forma a atingir os resultados esperados;
3. Verificar se há coerência na ordem das perguntas e se estão dispostas da melhor forma;
4. Verificar se os termos utilizados nas perguntas são de fácil compreensão para qualquer classe de respondentes;
5. Determinar a necessidade de adicionar perguntas específicas ou eliminar questões que se mostraram irrelevantes;
6. Averiguar a adequação das instruções para os entrevistadores.

A detecção antecipada das situações como as enumeradas acima evita custos desnecessários tanto em termos de tempo e dinheiro quanto em termos da credibilidade dos institutos de pesquisa. Não há uma quantidade recomendada de testes, porém os mesmos devem ser repetidos até que o questionário esteja realmente pronto para ser aplicado à população de estudo. Quando os questionários são desenvolvidos segundo as recomendações elencadas no Capítulo 2, geralmente dois ou três testes são suficientes para sua validação.

O número necessário de respondentes para a realização dos testes varia de acordo com seus objetivos, com a recomendação de amostras com cerca de 50 entrevistados para detectar as principais falhas, dificuldades e fraquezas de um questionário preliminar (PRESSER et al., 2004). Porém, existem estudos e pesquisas que visam captar diferenças nas respostas entre grupos por existência de uma ou mais características de interesse da investigação; ou, ainda, estudos em que haja necessidade de verificar se há diferença na

compreensão das perguntas por características demográficas ou recortes espaciais. Nesses casos, recomenda-se a seleção de amostras com tamanho suficiente para a estimação de parâmetros estatísticos e que cada um desses grupos ou características esteja bem representado na amostra selecionada.

Em relação aos resultados dos testes, recomenda-se que seja elaborado um relatório de validação do questionário, contendo informações que foram relevantes para a revisão efetuada, tais como: número de testes realizados; sugestões construtivas e concretas surgidas em cada rodada; alternativas de perguntas ou categorias de resposta revisadas ou abandonadas; reação dos respondentes vis-à-vis qualidade das respostas obtidas; reação dos entrevistadores e as dificuldades encontradas na obtenção das respostas; pertinência de cada questão do ponto de vista dos dados obtidos e da tabulação planejada. Recomenda-se ainda que essa documentação faça parte dos metadados da pesquisa.

Existem duas categorias básicas de métodos para testes relacionados às pesquisas: pré-campo e campo. Os métodos de pré-campo são utilizados nos primeiros estágios da elaboração de um instrumento de coleta e envolvem as atividades ligadas ao planejamento do que se pretende mensurar, formulação das perguntas adequadas para se obter as informações desejadas e definição da ordem das perguntas e do aspecto visual.

Os métodos de campo são utilizados quando não se pode avaliar o sucesso dos componentes do processo de pesquisa a priori, ou seja, quando não é possível testá-los antes da implementação da pesquisa como um todo. Deve-se projetar os testes de campo de forma a refletir as condições reais da pesquisa, incluindo aquelas passíveis de apresentar dificuldades operacionais, como seleção das amostras, coleta de dados e recursos eletrônicos relacionados, especificações de crítica, processamento de dados, estimativas, criação de arquivos e tabulações. Embora cada um desses aspectos da pesquisa possa ser testado separadamente, em caso de pesquisas de grande porte, cujos resultados sejam de maior relevância para o planejamento de políticas públicas e privadas, faz-se necessário um teste completo de todos os componentes, em uma espécie de ensaio geral. Cabe mencionar que os testes de campo também podem ser utilizados na avaliação de instrumentos de coleta.

Para fins do presente texto, optou-se por restringir o escopo aos métodos de pré-campo para avaliação e aprimoramento de questionários formais padronizados como instrumento de coleta, que se caracterizam por uma redação específica das perguntas e suas respectivas categorias de resposta, quando aplicáveis, além de uma ordem pré-determinada dos temas a serem pesquisados.

Métodos para testes pré-campo

Os testes pré-campo permitem avaliar o grau de compreensão das perguntas de um questionário, as dificuldades na elaboração das respostas e quaisquer outros eventos ocorridos durante as entrevistas ou preenchimento do questionário que possam introduzir erros de medição. Seus resultados permitem também avaliar quão bem as respostas obtidas representam o fenômeno que se está buscando medir, seja comparando os resultados com fontes externas ou edições anteriores da mesma pesquisa, o que se relaciona com questões

de validação ou viés de resposta, seja repetindo a aplicação do teste, o que aborda questões de confiabilidade ou variação das respostas obtidas. Dessa forma, os testes pré-campo se constituem em uma espécie de certificação de padrões de qualidade que as perguntas de um questionário devem atingir. Os padrões de qualidade de um questionário são: conteúdo (as perguntas elaboradas são efetivamente referentes ao tema que se pretende mensurar); cognição (os respondentes entendem facilmente as perguntas; os respondentes possuem a informação necessária para respondê-las e estão dispostos a responder); e usabilidade (facilidade tanto dos entrevistadores quanto dos respondentes em completar os questionários). Cabe mencionar que um questionário validado deve satisfazer simultaneamente aos requisitos dos três padrões.

A seguir, apresentam-se métodos reconhecidos por especialistas em pesquisa para verificar se o questionário atingiu cada padrão de qualidade supramencionado. Cada método produzirá um tipo de informação relevante para os padrões de qualidade, portanto recomenda-se que sejam elaborados testes que permitam avaliar mais de um padrão ao mesmo tempo. Os métodos de teste pré-campo são:

1. Revisões metodológicas, em que catedráticos no tema e especialistas em desenho de questionário analisam as perguntas para avaliar se sua forma e conteúdo são apropriados para medir os conceitos pretendidos. Uma vez que esse método não envolve diretamente o respondente, seus resultados devem ser complementados por outro método para atender aos padrões de qualidade.
2. Grupos focais, em que os pesquisadores realizam discussões semiestruturadas (focadas) com grupos da população-alvo para explorar seu conhecimento das questões relacionadas aos temas abrangidos pelo questionário. Nessas discussões, deve-se abordar como os especialistas pensam sobre as questões formuladas e se os termos utilizados são os mais adequados para falar sobre as mesmas. Embora envolva os respondentes, esse método deve ser complementado por outro método de teste para a validação do questionário como um todo, a menos que os participantes do grupo focal sejam entrevistados ou preencham e discutam um questionário autoadministrado.
3. Entrevistas cognitivas, nas quais os entrevistadores aplicam as perguntas elaboradas em entrevistas individuais, registrando como os entrevistados entendem as perguntas e formulam suas respostas, assim como qualquer outra condição que possa interferir na entrevista, como desconforto, interrupções, entre outros. Esse método pode ser utilizado em conjunto com outros métodos para avaliar se o questionário atende aos três padrões de qualidade.
4. Testes de usabilidade, que avaliam o desempenho do questionário em relação à facilidade de utilização pelos entrevistadores e respondentes, em condições controladas similares àquelas nas quais o questionário será aplicado. Durante o teste, os participantes tentam preencher o questionário ou realizar a entrevista enquanto os observadores assistem, ouvem e tomam notas com o objetivo de identificar quaisquer problemas de usabilidade, coletar dados qualitativos e quantitativos, e determinar a satisfação do participante com o produto. Esse método

pode ser utilizado para avaliar se o questionário atende aos três padrões de qualidade.

5. Experimentos Aleatorizados, nos quais grupos distintos da amostra teste recebem questionários contendo diferentes formulações de perguntas sobre o mesmo tema.

De acordo com a literatura, recomenda-se combinar dois ou mais métodos de avaliação do questionário em um mesmo teste pré-campo, denominado teste piloto de questionário. Nesse teste, os entrevistadores realizam as entrevistas usando procedimentos de amostragem e de campo semelhantes à pesquisa de fato. Sugere-se que o número de entrevistas seja próximo a 100 (GROVES et al., 2009; WEINBERG apud ZUKERBERG, 1995) ou estatisticamente significativo caso haja necessidade de avaliar diferenças por recorte espacial ou populacional. Através das observações de campo dos entrevistadores no teste piloto de questionário, é possível detectar problemas em perguntas difíceis de ler da forma que estão escritas ou que sejam difíceis de responder. Além disso, os dados obtidos podem ser tabulados e analisados para verificar problemas como taxa de não resposta ou viés, assim como ajudar a compreender o significado das respostas da pesquisa.

Por fim, é importante lembrar que a inclusão de qualquer tema novo ou reformulação de temas previamente investigados sempre requer testes prévios para avaliação das perguntas, visando assegurar uma coleta bem-sucedida e a produção de resultados confiáveis.

Revisões Metodológicas

Revisões Metodológicas são um método de teste pré-campo de questionário de baixo custo, curta duração e fácil implementação, que deve ser realizado na fase inicial do processo de desenvolvimento do questionário, e consiste na avaliação do questionário proposto por um grupo de especialistas, buscando identificar possíveis problemas tanto para entrevistadores quanto para respondentes. O grupo de especialistas deve ser composto por metodologistas de pesquisa ou especialistas em desenho de questionários e catedráticos no tema da pesquisa, com experiência reconhecida em avaliação de questionários. Embora esse grupo tenda a ser pequeno, sendo recomendado um mínimo de três revisores (OLSON, 2010), é possível haver um grupo com mais de 20 revisores (WILLIS et al., 1999), dependendo da complexidade do tema proposto.

Através das Revisões Metodológicas, é possível identificar estruturas linguísticas problemáticas, ordem das perguntas ou blocos temáticos que levem a quebras de raciocínio no processo de compreensão da pergunta e elaboração da resposta, discernir formulação das perguntas e das categorias de resposta que sejam propensas a não serem respondidas ou que gerem imprecisão na resposta, além de classificar as perguntas de acordo com sua propensão a erros de medição (OLSON, 2010). O resultado dessas revisões deve ser apresentado por cada especialista em forma de relatório.

Pode-se conduzir o grupo de especialistas de duas formas, seja através de uma discussão estruturada sobre cada pergunta, com um esquema padronizado de codificação,

seja através de uma discussão não estruturada sobre cada pergunta e sem uma lista padronizada de codificação. Em ambas, há a presença de um moderador. Durante a discussão, recomenda-se avaliar as características motivacionais e cognitivas das questões, pontuando se a pergunta ou característica medida pela questão requer do respondente um grande esforço cognitivo; se o tema é sensível, ou seja, se envolve informações constrangedoras ou particulares; ou ainda se a pergunta pode revelar um comportamento socialmente indesejável. Recomenda-se avaliar, ainda, se a pergunta é passível de proporcionar falhas no processo de elaboração da resposta.

Embora seja um método de fácil implementação, é um método sujeito a interpretações e julgamentos particulares dos revisores. Dessa forma, pode haver falhas nesse método como o fato de diferentes especialistas não identificarem os mesmos problemas em um questionário ou não examinarem uma determinada pergunta com a devida atenção. Cabe mencionar ainda que as Revisões Metodológicas são o único método de teste que não envolve os respondentes e, portanto, para a efetiva validação do questionário, deve-se aplicar algum outro método complementar de teste pré-campo.

Grupos Focais

Os Grupos Focais são grupos de debate sobre o tema da pesquisa que ocorre entre os pesquisadores e um número pequeno de voluntários da população-alvo, sob a supervisão de um moderador. Durante a discussão, os voluntários são encorajados a expressar suas opiniões sobre as perguntas, inclusive concordando ou discordando dos demais membros do debate ou influenciando outros membros com seus pontos de vista. O objetivo dos Grupos Focais é auxiliar os pesquisadores a entender os conceitos e sua nomenclatura, além das perspectivas comuns e principais pontos sobre o tema do ponto de vista da população-alvo.

O debate deve ser conduzido pelo moderador de forma a criar uma atmosfera leve, participativa e colaborativa, envolvendo o grupo na discussão e criando transições leves entre as perguntas a serem analisadas. Recomenda-se que esses debates ocorram em um ambiente especialmente preparado, que permita observar ou gravar sem constranger os participantes. O produto do Grupo Focal pode ser um relatório contendo observações pontuais para cada pergunta, um resumo das principais observações ou mesmo uma transcrição completa do debate.

Em termos de contribuição para o aprimoramento do questionário, os Grupos Focais permitem aos pesquisadores identificar e compreender os termos utilizados pela população-alvo, e de que forma essa população percebe os tópicos explorados pelo questionário como relevantes ou não ao tema. Também auxiliam a determinar como os respondentes organizam suas ideias e opiniões sobre o tema em estudo, além de permitir que os pesquisadores entendam as experiências vivenciadas pelos participantes em relação ao tema, o que permite uma melhor adequação das perguntas do questionário.

Apesar de ser um método pré-campo de baixo custo e rápida duração, os Grupos Focais têm suas limitações. Os participantes podem não ser representativos da população-alvo da pesquisa, tornando as observações do debate apenas percepções ou experiências

peçoais daquele grupo em particular, sem possibilidade de generalização; além disso, as informações obtidas são majoritariamente qualitativas e não permitem que quaisquer resultados ou conclusões sejam replicadas. Da mesma forma, quando o debate é conduzido de forma a somente apresentar o tema como objeto das discussões ou sem a aplicação do questionário aos voluntários, simulando uma entrevista, torna-se difícil uma real avaliação das dificuldades motivacionais e cognitivas associadas às perguntas e suas categorias de resposta. Não obstante, esse método é eficiente quando utilizado em conjunto com outros métodos de teste pré-campo no início do processo de elaboração do questionário.

Entrevistas Cognitivas

As Entrevistas Cognitivas permitem avaliar o quanto uma pergunta ou um conjunto de perguntas de uma pesquisa capta a informação que se necessita saber com a precisão necessária. Esse método consiste na realização de entrevistas semiestruturadas conduzidas com grande nível de detalhamento em uma amostra pequena de informantes, seja em ambientes controlados como laboratórios, seja em outros ambientes propícios que permitam gravação ou observação para análise posterior. Essas entrevistas devem ser conduzidas por pesquisadores com experiência em desenvolvimento de questionário ou entrevistadores com experiência em avaliação de perguntas e sua realização deve ocorrer após a construção ou modificação do questionário proposto com base nos relatórios dos Grupos Focais e das Revisões Metodológicas (EUROSTAT, 2006).

O nome do método vem do fato que, embora quase instantâneo, o processo de responder a uma pergunta em um questionário não é simples e envolve uma série de processos cognitivos. Esses processos se dividem em quatro estágios: compreensão, busca na memória, julgamento e resposta, cuja definição e os erros associados encontram-se no Quadro 2.

Quadro 2. Definições dos estágios do processo cognitivo de elaboração de respostas e erros associados.

Estágios Cognitivos	Definição	Causas de possíveis erros de resposta
1 – Compreensão	O informante interpreta a pergunta	Termos desconhecidos, conceitos ambíguos, longos e complexos
2 – Busca na memória	O informante busca na memória informações relevantes	Dificuldade de lembrança
3 – Julgamento	O informante avalia ou calcula a resposta	Pergunta enviesada ou tema sensível. Dificuldade de avaliação
4- Resposta	O informante fornece a informação no formato solicitado	Opções de resposta incompletas

Nas Entrevistas Cognitivas, verifica-se o desempenho das perguntas examinando o processo pergunta-resposta e considerando o grau de dificuldade experimentado pelo respondente para compreender a pergunta e formular uma resposta precisa. Ou seja, é possível investigar a forma com a qual as pessoas entendem, processam mentalmente e respondem a uma determinada pergunta, através de técnicas que permitem concluir se a redação da pergunta ou das opções de resposta, assim como as instruções, são interpretadas

da forma que se pretende. Entre seus objetivos, destacam-se a verificação da existência de um entendimento comum entre o pesquisador e o informante, de forma a gerar dados válidos; a existência de barreiras ocultas que dificultam o entendimento das perguntas; e, se há algum tema que os informantes considerem de entendimento difícil ou que levem a algum equívoco e que possa ser corrigido antes do trabalho de campo, melhorando a qualidade dos dados da pesquisa.

Existem diversas técnicas utilizadas na condução das Entrevistas Cognitivas, que permitem, inclusive, a combinação entre elas para obter resultados mais precisos. As principais técnicas estão descritas a seguir:

1. Protocolo Verbal Pensar em Voz Alta (Think-a-loud), que consiste em que o respondente fale em voz alta todos os pensamentos que o levaram a responder daquela forma. Esse protocolo é utilizado na identificação de dificuldades na compreensão das perguntas e de percepções equivocadas da tarefa de responder às perguntas, além de evidenciar os tipos de estratégia utilizada pelo respondente na recuperação da memória e as reações esboçadas frente aos temas sensíveis. Esse protocolo pode ser aplicado durante ou após a formulação da resposta e sua eficácia depende da capacidade e disposição do respondente em articular seus pensamentos.
2. Sondagem (Probing), que consiste na elaboração de perguntas de sondagem, com elementos relevantes para a compreensão da pergunta original e que possam ser respondidas livremente ou através de categorias fixas de resposta, captando o processo cognitivo pelo qual a pessoa elaborou a resposta à pergunta original.
3. Paráfrases, em que os respondentes devem repetir as perguntas com suas próprias palavras. Dessa forma, é possível examinar a interpretação e o grau de compreensão da pergunta pelo respondente, assim como a concordância com o conceito que se pretende medir através da mesma. A técnica de parafrasear também permite obter melhores formulações para perguntas, quando, por exemplo, respondentes distintos utilizam consistentemente a mesma terminologia. Existem duas abordagens principais para essa técnica, sendo a primeira quando a pergunta deve ser repetida palavra por palavra e a segunda, quando o contexto da pergunta é repetido nas palavras do respondente, sendo essa última considerada mais eficiente. Por fim, a Paráfrase é especialmente útil na detecção de perguntas ou categorias de resposta que possam ser confusas.
4. Avaliações de confiança, em que o respondente é solicitado a avaliar o grau de confiabilidade de suas respostas, classificando-as com a ajuda de uma determinada escala, o que permite identificar perguntas difíceis de responder. Entende-se que, em geral, baixas classificações de confiança resultam da falta de conhecimento sobre o tema ou da dificuldade de lembrança.
5. Latência de resposta, que é o lapso de tempo transcorrido entre a pergunta e a resposta, medido por um cronômetro ou por estimativa. Utilizam-se os métodos de latência de resposta na identificação de problemas de interpretação das

perguntas, recuperação de memória e seleção de respostas. Essa técnica se aplica com maior facilidade às pesquisas pela internet ou assistidas por computador.

Ou seja, a ênfase do método de Entrevistas Cognitivas não está na resposta específica em si, mas no entendimento do processo de cognição dos informantes, isto é, na forma com a qual os informantes interpretam as perguntas e porque eles deram aquela resposta em particular, verificando o que veio à mente do informante ao ouvir a pergunta, pedindo para que o informante elabore a mesma pergunta em suas próprias palavras e verificando tanto a dificuldade da pergunta quanto se o leiaute da pergunta facilitou ou dificultou o entendimento, entre outros. Cabe ainda mencionar que esse método não se restringe aos testes de perguntas elaboradas para questionários, podendo ser aplicado para testar outros documentos, como formulários¹ autoadministrados ou administrados pelo entrevistador. Recomenda-se que os resultados e aprimoramentos implementados a partir desse método sejam apresentados em forma de relatório, devendo constar dos metadados da pesquisa.

Testes de Usabilidade

Os Testes de Usabilidade são utilizados para abordar questões relacionadas à facilidade com que os usuários - entrevistadores e respondentes – utilizam todo o sistema da pesquisa, inclusive o questionário, que pode estar disponibilizado em papel, dispositivo móvel de coleta ou pela internet. Nesses testes, o foco está na avaliação do usuário sobre questões ligadas ao software, hardware, leiaute de tela apropriado em relação ao método de coleta de dados, legibilidade das perguntas, duração da entrevista e clareza das instruções, seja no manual, visualizadas pelo computador ou dadas durante o treinamento.

Existem dois tipos de Testes de Usabilidade, sendo o primeiro composto por métodos de inspeção de usabilidade e, o segundo, por métodos de avaliação dos usuários finais. Nos métodos de inspeção de usabilidade, também conhecidos na literatura como avaliações heurísticas, um grupo de especialistas avalia o questionário com base em um conjunto de heurísticas ou critérios de avaliação (EUROSTAT, 2006). Esses critérios de avaliação podem ser a facilidade de navegação ou de transição entre os blocos temáticos, a consistência do questionário e a minimização da carga sobre o usuário. A inspeção de usabilidade tem as mesmas características que o método de Revisões Metodológicas e segue as mesmas recomendações.

Já o método de avaliação dos usuários finais envolve diretamente os entrevistadores e respondentes, podendo ser realizado em ambientes controlados ou em simulações de entrevistas no campo, em bases experimentais ou não-experimentais. A vantagem de realizar o Teste de Usabilidade em ambientes controlados é a possibilidade de gravação do teste ou observação direta da tela do entrevistador através de projeção da mesma. Porém, em campo,

¹ O formulário é um dos instrumentos essenciais para a investigação social, cujo sistema de coleta de dados consiste em obter informações diretamente do entrevistado. Segundo Sellitz apud Oliveira et al. (2016), formulário "é o nome geral usado para designar uma coleção de questões que são perguntadas e anotadas por um entrevistador numa situação face a face com outra pessoa". Não obstante, para fins do presente texto, respeitou-se a nomenclatura utilizada pelo IBGE em relação ao conceito acima (questionário).

há a vantagem de testar as perguntas no próprio ambiente da pesquisa, longe da configuração artificial de um laboratório.

Através dos Testes de Usabilidade, pode-se obter uma maior confiança de que a forma com que as perguntas são feitas e respondidas está de acordo com seu objetivo original. Além disso, os testes auxiliam na identificação de problemas de natureza diferente, como questões e críticas surgidas durante o treinamento de entrevistadores. Embora não haja uma fase específica para sua realização, recomenda-se que esse método seja aplicado no estágio inicial de desenvolvimento do sistema de suporte ao questionário, para que seus resultados sejam utilizados no aprimoramento do mesmo (EUROSTAT, 2006). Recomenda-se que os resultados e aprimoramentos implementados a partir desse método sejam apresentados em forma de relatório e constem dos metadados da pesquisa.

Experimentos Aleatorizados

Os Experimentos Aleatorizados permitem determinar os efeitos de medição e comparar os desempenhos de diferentes versões de perguntas sobre um mesmo tema, de questionários ou instrumentos de coleta distintos, ou ainda de diferentes procedimentos ou instruções de campo, através de sua aplicação em subamostras selecionadas aleatoriamente. Cabe destacar que o tamanho das subamostras deve ser suficiente para permitir que se possa medir diferenças significativas. Outrossim, cabe reforçar a necessidade de selecionar os respondentes de forma aleatória para aplicar cada tipo de questionário ou procedimento, para que as diferenças encontradas possam ser efetivamente atribuídas à pergunta ou questionário e não aos efeitos de amostras incomparáveis (EUROSTAT, 2006).

O método de Experimentos Aleatorizados possibilita evidenciar de forma clara os impactos das características metodológicas como diferenças na formulação, linguagem ou ordem das perguntas, assim como do método de coleta de dados sobre as respostas obtidas. Destaca-se ainda como vantagem, o controle do viés entre as subamostras selecionadas aleatoriamente, permitindo uma descrição causal dos efeitos, ou seja, das consequências que podem ser atribuídas à variação deliberada na metodologia da pesquisa. Esse método também pode ser utilizado para calibrar diferenças ou conectar a versão original de uma pergunta ou bloco temático e as versões aprimoradas em pesquisas longitudinais ou sazonais, permitindo aos pesquisadores conduzir as análises de tendência.

Não obstante, esse método não permite concluir qual versão dos questionários ou procedimentos produz dados que reflitam melhor a realidade do tema pesquisado, a não ser que sejam coletados dados de validação oriundos de fontes externas, contra os quais as respostas da pesquisa possam ser verificadas ou quando existem fortes razões teóricas para decidir que uma versão das perguntas é melhor que outra (GROVES et al., 2009). Outra limitação do método reside no fato de que a realização de Experimentos Aleatorizados costuma ter um custo alto e demandar maior tempo do que os demais métodos, em especial quando conduzida sob condições de pesquisa de produção. Além disso, os Experimentos Aleatorizados podem sofrer limitações semelhantes aos demais métodos na avaliação dos dados da pesquisa, isto é, a inferência para uma população-alvo só é possível se uma amostra

estatística com poder estatístico suficiente for extraída da população-alvo para conduzir o experimento.

Recomenda-se a realização de Experimentos Aleatorizados durante o planejamento da pesquisa, em quaisquer testes de questionário previstos, ou mesmo durante a coleta da pesquisa, quando o objetivo principal é a produção de estimativas visando a avaliação da qualidade dos dados e o objetivo secundário, a avaliação do impacto de perguntas, instruções ou procedimentos diferentes. Os resultados e aprimoramentos implementados a partir desse método devem ser apresentados em forma de relatório, que deve constar dos metadados da pesquisa.

Confiabilidade e Validade de um instrumento de coleta

Sabe-se a priori que as principais causas de variação entre indicadores oriundos de diferentes bases de dados são as diferenças quanto aos critérios metodológicos adotados por cada pesquisa, assim como a forma de obtenção das informações, em especial, a diferença entre os instrumentos de coleta. Em face da existência de diversos instrumentos elaborados para captar uma determinada informação, a escolha por um deles deve passar por uma avaliação prévia quanto à sua confiabilidade e sua validade.

Segundo Martins (2006), a confiabilidade e a validade são requisitos de toda medida ou indicador. Um indicador confiável é consistente e passível de replicação, sendo uma representação válida e precisa do que se pretende medir. No caso de um instrumento de coleta, a confiabilidade pode ser determinada tanto através da constância dos resultados quanto da comparação dos resultados após sucessivas aplicações ao mesmo indivíduo, objeto ou grupo. Ou seja, entende-se por confiabilidade a consistência ou estabilidade de uma medida ou indicador (SELLTIZ, WRIGHTSMAN e COOK, 1987). Havendo flutuação entre as medições, deve-se avaliar se essa flutuação resulta de erros de mensuração ou de diferenças reais entre os indicadores, o que pode ser dado pelo desvio padrão: quanto menor o desvio padrão, maior a confiabilidade do instrumento de coleta. Pode-se aplicar ainda a metodologia de testes de hipótese para verificar a existência de diferenças significativas entre os indicadores.

Destacam-se ainda outras técnicas para avaliar a confiabilidade de um instrumento de coleta: teste-reteste, formas equivalentes, metades partidas e avaliadores. A técnica do teste-reteste consiste na aplicação do instrumento de coleta de dados duas vezes a um mesmo grupo de pessoas e, verificando-se uma correlação fortemente positiva entre os resultados, pode-se concluir que o instrumento é confiável. Na técnica de formas equivalentes, são aplicadas duas versões similares do instrumento de coleta ao mesmo grupo de indivíduos dentro de um curto período de tempo, avaliando-se a variação dos padrões de respostas. A técnica das metades partidas requer uma única aplicação do instrumento de coleta com questões equivalentes em conteúdo e dificuldade, dividindo-se posteriormente os conjuntos

de questões e comparando seus resultados; caso esses resultados estejam fortemente correlacionados, o instrumento é considerado confiável. No caso da técnica de avaliadores, dois ou mais avaliadores passam por um mesmo treinamento e seguem as mesmas instruções na aplicação do instrumento de coleta; caso valores similares sejam observados pelos diferentes avaliadores, pode-se dizer que o instrumento de coleta é confiável.

Pode-se definir a validade de um instrumento de coleta como a sua capacidade em realmente medir a variável que pretende medir (KELSEY et al., 1996). Ou seja, um instrumento válido para medir alguma característica deve realmente medir essa característica e não alguma outra que possa influenciar na medição como, por exemplo, investigar a capacidade de leitura vis-à-vis conhecimento prévio de um texto pelo respondente. A validade pode ser avaliada através das evidências de validade de conteúdo, validade de critério e validade de construção.

A validade de conteúdo se refere ao grau em que o instrumento abrange os diferentes aspectos do objeto ou tema que pretende medir, não contendo elementos que possam ser atribuídos a outros objetos, ou seja, o instrumento de coleta deve ser exaustivo quanto às variáveis que compõem seu objeto de estudo. Para validar um instrumento de coleta segundo o critério, estabelece-se anteriormente um critério considerado padrão e se avalia estatisticamente o quanto os resultados se identificam ou se afastam desse critério. E, finalmente, a validade de construção faz referência ao grau em que os indicadores obtidos através de um instrumento de coleta se relacionem com outros indicadores semelhantes, oriundos de bases de dados com o mesmo arcabouço teórico e conceitos estabelecidos. Para tal, realiza-se uma série de estudos que visam à verificação empírica, por meio de testes estatísticos, das construções teóricas sobre a relação entre as variáveis a serem medidas. Cabe ressaltar que a realização desses estudos está condicionada a uma amostra grande o suficiente para obter resultados válidos.

Recomendações gerais para realização de testes de questionário

Apresentam-se, a seguir, algumas recomendações extraídas da revisão bibliográfica para auxiliar na elaboração e avaliação das perguntas de um questionário. Destaca-se na literatura, como principal ponto, que todo o processo de elaboração do questionário seja norteado por uma estratégia consistente, adaptada para as condições da pesquisa e que permita detectar e eliminar o maior número possível de problemas através dos testes pré-campo. Entende-se aqui por estratégia a sequência de tarefas e métodos a serem implementados, considerando a influência de diversos fatores e circunstâncias particulares de cada pesquisa, de forma a tornar iterativo o processo de avaliação da qualidade das perguntas.

A ordem das tarefas e a escolha dos métodos a serem aplicados dependem principalmente do fato de se estar implementando uma nova pesquisa ou avaliando uma pesquisa em andamento. Portanto, visando auxiliar na escolha dos métodos mais adequados, optou-se por dividir as recomendações para cada etapa do ciclo de desenvolvimento do questionário por pesquisas novas ou em andamento. Segundo o Handbook of Recommended

Practices for Questionnaire Development and Testing in the European Statistical System (EUROSTAT, 2006), é altamente recomendável que cada uma das etapas de planejamento identificadas seja conduzida de forma ordenada. Em ambos os casos, recomenda-se ainda que todas as fases desse processo sejam exaustivamente documentadas e disponibilizadas como parte dos metadados da pesquisa.

Pesquisas novas

Em relação às pesquisas novas, a estratégia adotada para o processo de elaboração e avaliação do questionário deve cobrir cinco etapas, apresentadas na Figura 1. Ressalta-se que essas etapas estão de acordo com o Modelo Genérico do Processo de Produção Estatística (GSBPM). As recomendações para cada etapa encontram-se descritas a seguir.

Figura 1. Etapas da elaboração de testes de questionário para novas pesquisas



Fonte: Adaptado de EUROSTAT, 2006.

Recomenda-se que a etapa de **Especificação dos Conceitos** tenha início com uma revisão de literatura sobre o tema a ser investigado, inclusive com a leitura de relatórios sobre testes de questionários já realizados em pesquisas semelhantes, que possam ser comparáveis. Após a revisão da literatura, deve-se proceder à descrição das definições, conceitos e objetivos da pesquisa, com clareza e em conformidade com os modelos conceituais dos respondentes. Recomenda-se que a elaboração do texto de cada pergunta tenha como ponto de partida a revisão da literatura, sob a perspectiva tanto dos especialistas quanto dos entrevistadores e respondentes. A importância de estabelecer um bom marco conceitual impacta diretamente na qualidade das pesquisas novas. A população-alvo, o questionário, os recursos necessários para a realização da pesquisa e o cronograma são determinados a partir do marco conceitual. Além disso, o embasamento teórico deve permear todas as decisões relacionadas a quaisquer mudanças do questionário, determinando sua estrutura, lista de variáveis-alvo, plano tabular e método de coleta dos dados.

Nessa fase, as Revisões Metodológicas, os Grupos Focais e Entrevistas Cognitivas

são os métodos indicados, pois permitem que a complexidade dos conceitos teóricos e sua operacionalização em características empíricas sejam observadas, criando perguntas adequadas para representar o conceito temático.

O **Desenho de Questionário** advém da especificação da base conceitual, traduzindo as variáveis-alvo em perguntas concretas, estabelecendo sua sequência e buscando uma formulação apropriada, inclusive para as categorias de resposta. As recomendações básicas para construção de um questionário foram descritas anteriormente no Capítulo 2. Como recomendação adicional, destaca-se a observação dos efeitos de contexto resultantes da sequência dos blocos temáticos no questionário, considerando, inclusive, o método de coleta de dados. Em cada bloco temático, deve-se atentar para a necessidade de saltos, pois a introdução do dispositivo móvel de coleta e o preenchimento de questionários pela internet potencializam o uso de regras de salto, aumentando o nível de complexidade das perguntas e do questionário como um todo. Portanto, recomenda-se que a instrução esteja disponível e clara para o respondente quando necessário, especialmente em questionários autoadministrados. Recomenda-se ainda que, para entrevistas por telefone, as perguntas sejam curtas e diretas, com categorias de resposta reduzidas.

A elaboração das perguntas de um questionário também deve considerar a necessidade de comparações internacionais. Portanto, ao formular as perguntas, faz-se necessário também harmonizar definições e variáveis com aquelas adotadas internacionalmente e assegurar que as traduções dos questionários sejam tão precisas quanto possível. Segundo o EUROSTAT (2006), recomenda-se para essa harmonização:

1. Procurar na literatura internacional uma definição comum dos conceitos a serem medidos e garantir que essa definição seja realmente adotada por cada país, região ou cultura envolvidos na pesquisa;
2. Identificar indicadores válidos para as variáveis de interesse, considerando as especificidades de cada país, região ou cultura;
3. Decidir se o bloco temático de interesse deverá ser investigado pelo mesmo questionário (harmonização de insumos) ou pelos questionários específicos de cada país, região ou cultura, produzindo dados harmonizados após a coleta de dados (harmonização de resultados).
4. Testar se as perguntas do questionário refletem as estruturas empíricas encontradas nos diferentes países, regiões ou culturas e se estão logicamente relacionadas com a definição comum encontrada na literatura.

Neste caso, a harmonização do bloco temático deve ser feita de forma cuidadosa, garantindo que o questionário seja compreendido e respondido igualmente pelos entrevistados de diferentes contextos nacionais e culturais. Portanto, a tradução de perguntas que envolvam atitudes e comportamentos deve ser integrada em um processo de testes, revisões e decisão do texto final por especialistas nas diversas línguas e culturas, além da equipe de tradutores e especialistas temáticos. Cabe ressaltar a importância de realizar testes da tradução com métodos cognitivos e quantitativos, para verificar a existência de erros de classificação de itens ou variáveis por meio de procedimentos estatísticos e utilizando dados

comparativos dos países, regiões ou culturas envolvidas (EUROSTAT, 2006).

Finalizada a redação preliminar das perguntas, deve-se proceder à elaboração do desenho gráfico do questionário, considerando o método de coleta de dados para tornar o leiaute interessante e atrativo para entrevistadores e respondentes. Sabe-se que elementos gráficos ativam os níveis emocionais, funcionais e reflexivos dos seres humanos; portanto, a utilização de estímulos não verbais como símbolos e figuras pode guiar entrevistadores e respondentes no preenchimento do questionário, melhorando sua usabilidade e compreensão das perguntas, e propiciando a obtenção de estimativas válidas e confiáveis (EUROSTAT, 2006).

Para a etapa de **Teste de Questionário**, recomenda-se que os testes pré-campo sejam realizados a partir de uma primeira versão do questionário, que deve ser muito semelhante àquela que será usada para a coleta de dados real, tanto em termos do texto das perguntas como do desenho gráfico. Os testes de questionário devem ter como base três pontos essenciais: (1) estrutura, sequência e formulação de perguntas e respectivas categorias de resposta; (2) problemas relacionados à harmonização de conceitos e tradução; e (3) método de coleta de dados, considerando a utilização de entrevistadores ou questionários autoadministrados. Para abranger esses diferentes pontos, faz-se necessária uma combinação de diferentes métodos de teste de questionário, de forma que os padrões de qualidade sejam alcançados. A recomendação mínima é a realização de pelo menos um teste composto principalmente por respondentes em potencial e não apenas por especialistas.

Conforme visto na Seção 3.1, existem diversos métodos de testes pré-campo e sua adequação depende de vários fatores e circunstâncias, incluindo o tipo e tamanho da pesquisa, as características da população-alvo, o conteúdo da pesquisa, a utilização de questões-padrão previamente testadas e o método de coleta de dados. No caso de pesquisas novas, os testes de questionário devem ser mais intensivos e repetidos o quanto for necessário dentro do orçamento determinado para essa etapa da pesquisa. Recomenda-se que o questionário resultante dessas revisões seja testado em um estudo piloto de maior escala em termos de trabalho de campo. Por fim, deve-se tomar cuidado para garantir uma representação apropriada da população-alvo ao selecionar os respondentes com o propósito de testar questionários.

A etapa de **Revisão do Questionário** é o desdobramento da etapa de teste como um processo iterativo até que se alcance um questionário que possa ser considerado ideal. Os resultados dos testes realizados no estágio inicial de elaboração do questionário podem gerar revisões na formulação das perguntas ou categorias de resposta, na sequência lógica das perguntas ou outras alterações que devem ser testadas novamente para avaliar seus efeitos. Assim, no contexto da revisão do questionário testado, o objetivo é verificar se as mudanças propostas estão realmente resultando em maior validade e confiabilidade dos dados em relação aos objetivos específicos da pesquisa. Nessa etapa, podem ser utilizados diferentes métodos de teste ou ainda repetir o mesmo teste utilizado na etapa anterior.

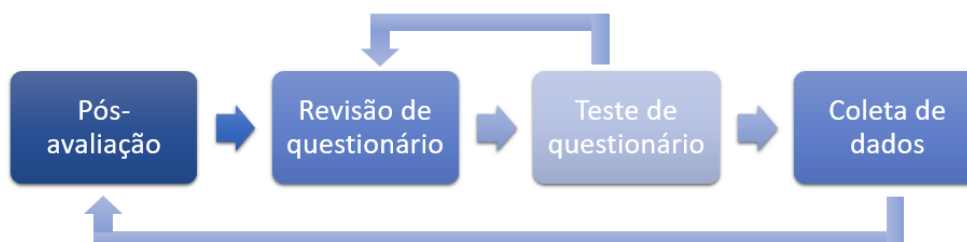
A etapa de **Coleta Final de Dados** encerra o processo iterativo de testes e revisão do questionário. Porém, recomenda-se que o processo de observação seja contínuo tanto através do monitoramento do desempenho dos entrevistadores e suas observações de campo

(paradados da pesquisa), quanto através de consulta aos respondentes. Esse monitoramento do trabalho de campo pode ser considerado como uma ferramenta contínua de avaliação e está em conformidade com os Princípios de Gestão da Qualidade dados pela ISO 9001:20152, tornando-se essencial para a avaliação posterior da pesquisa.

Pesquisas em andamento

A estratégia para testes dos questionários de pesquisas em andamento segue etapas ligeiramente diferentes em termos de ordem, além de partir da pós-avaliação da pesquisa. A Figura 2 ilustra as etapas de avaliação do questionário de pesquisas em andamento e a estratégia para cada etapa está descrita a seguir. Cabe mencionar que essas etapas também estão em conformidade com o Modelo Genérico do Processo de Produção Estatística (GSBPM).

Figura 2. Etapas da elaboração de testes de questionário para pesquisas em andamento



Fonte: Adaptado de EUROSTAT, 2006.

A etapa de **Pós-avaliação** das pesquisas em andamento tem início com a análise dos dados disponíveis, considerando ainda a análise das distribuições das respostas, inclusive por características sociodemográficas ou outras de interesse; taxas de não resposta de itens; crítica dos dados; taxas de imputação; análise dos resultados inconsistentes; e validação externa. Caso seja encontrado algum aspecto implausível, os relatórios de campo e de consulta aos respondentes podem auxiliar no esclarecimento, identificando perguntas ou categorias de resposta que apresentaram problemas. Havendo disponibilidade de dados e relatórios de campo, essa verificação pode ser realizada durante o processo de coleta da pesquisa. Recomenda-se que a pós-avaliação seja realizada pela equipe responsável pela

² Os sete princípios de gestão da qualidade dados pela ISO 9001:2015 são (1) Foco no Cliente; (2) Liderança; (3) Engajamento da equipe; (4) Abordagem por processos; (5) Melhoria contínua; (6) Tomada de decisões com base em evidências; e (7) Gestão de relacionamentos.

pesquisa ou bloco temático em conjunto com a equipe de especialistas em testes de questionário, e seus resultados devem estar documentados nos metadados da pesquisa. Caso os resultados da pós-avaliação detectem problemas, recomenda-se proceder às etapas de Revisão e Teste de Questionário descritas na Seção 3.3.1.

Em pesquisas contínuas, pode-se aplicar a estratégia de reduzir a avaliação às perguntas que apresentaram problemas, restringindo o teste a algumas perguntas selecionadas ou blocos temáticos do questionário. Porém, a realização de testes é sempre proveitosa, mesmo que não haja algum problema óbvio. Portanto, recomenda-se uma revisão periódica dos blocos temáticos para verificar se as perguntas formuladas e as respectivas categorias de resposta ainda estão em conformidade com os conceitos que se pretende medir. Nesse caso, cabe ressaltar que há limitações a serem consideradas, pois quaisquer revisões propostas – mesmo aparentemente menores – podem surtir efeitos sobre a continuidade da série histórica e seus resultados devem ser observados cuidadosamente, necessitando de um teste piloto de questionário antes da coleta de dados propriamente dita.

Questionário do próprio

Em princípio, supõe-se que informações prestadas pelo respondente sobre si mesmo sejam mais precisas do que por terceiros. Entretanto, na prática, nem sempre é possível obter respostas de cada indivíduo, seja pela ausência ou impossibilidade (TUCKER E MILLER, 1993). Nesse caso, um ou mais respondentes fornecem as informações para os indivíduos impossibilitados ou ausentes por ocasião da entrevista, sendo denominados informantes-chave, secundários, substitutos ou *proxy* (na língua portuguesa, “procurador” ou “representante”). Essas situações são comuns em pesquisas domiciliares e censos, em que, geralmente, o mesmo respondente presta as informações para todos os demais membros do domicílio.

Segundo Moore (1988), a literatura não fornece evidências suficientes para afirmar uma maior precisão das respostas do próprio em relação às respostas fornecidas pelo *proxy*; porém, deve-se considerar que existem poucos estudos bem elaborados sobre essa questão, assim como os tópicos abordados são limitados. Nesses estudos, as diferenças entre as respostas do próprio e do *proxy* parecem estar relacionadas ao tipo da informação perguntada, à importância da informação para o próprio, ao tipo e grau de relacionamento entre próprio e *proxy*, e o método de aplicação da pesquisa. Ou seja, face à complexidade dos diversos arranjos domiciliares, a precisão dos dados fornecidos depende da medida em que o *proxy* tem conhecimento detalhado ou familiaridade com as características e particularidades de todos os indivíduos pertencentes à unidade investigada. Cabe mencionar, ainda, a recomendação do EUROSTAT sobre não utilizar o *proxy* nos casos em que a pergunta é muito subjetiva, o tema é muito sensível ou o caráter das informações é notoriamente de menor conhecimento pelo *proxy* (EUROSTAT, 2018).

Como parte de uma estratégia abrangente dos testes de questionário, existem algumas recomendações a serem seguidas sobre temas que possam gerar imprecisão caso a informação seja prestada pelo *proxy*, como características sociodemográficas (cor, etnia ou

raça; fecundidade; migração; religião; rendimento) ou estados de saúde (cobertura vacinal, morbidade, pessoas com deficiência, uso de serviços de saúde), entre outros. Para esses temas, a combinação do método de Experimentos Aleatorizados com as Entrevistas Cognitivas é a mais recomendada, onde uma pessoa do domicílio é selecionada para responder o que se convencionou chamar de Questionário do Próprio (*self-reporting*). O Questionário do Próprio é uma réplica do questionário proposto, que deve ser aplicado exclusivamente à pessoa do domicílio com as características sociodemográficas ou estado de saúde que se deseja pesquisar. Já o questionário proposto deve ser aplicado a todos os indivíduos da unidade investigada, inclusive à pessoa selecionada para responder o Questionário do Próprio. As respostas obtidas devem ser comparadas, em especial aquelas relacionadas às perguntas de sondagem. Dessa forma, é possível obter as diferenças entre as percepções do *proxy* e do próprio em relação às características investigadas.

Testes de questionário realizados no IBGE

Em termos de testes de questionário realizados pelo IBGE, destacam-se os testes relacionados aos levantamentos censitários, com o objetivo de testar tanto novas propostas metodológicas de investigação do tema quanto melhorar a precisão das informações em um determinado bloco temático. Esses testes estão listados a seguir. Para maiores detalhes, pede-se consultar os respectivos relatórios.

1. Bloco temático sobre Pessoas com Deficiência, realizado por ocasião dos Censos Demográficos 2000 e 2010, para verificar a adequação de uma nova proposta metodológica de investigação do tema, segundo recomendações internacionais. O método de teste pré-campo utilizado foi uma combinação das Entrevistas Cognitivas com uma adaptação dos Experimentos Aleatorizados;
2. Bloco temático sobre Migração, realizado por ocasião do Censo Demográfico 2010, para testar perguntas sobre os temas Migração Interna, Internacional e Deslocamento, aqui definido como a migração pendular diária para trabalho, estudo, ou outros serviços. O método de teste pré-campo utilizado foi uma combinação de Teste de Usabilidade, Entrevistas Cognitivas com uma adaptação dos Experimentos Aleatorizados;
3. Tema Alfabetização, realizado por ocasião do Censo Demográfico 2010, para investigar a influência de inserir a instrução sobre ler e escrever um bilhete simples no texto da pergunta sobre alfabetização. Como instrução de trabalho, o manual do recenseador explicava que deveriam responder afirmativamente à pergunta aquelas pessoas que se consideravam capazes de escrever pelo menos um bilhete simples no idioma que conheciam. O método utilizado foi uma combinação das Entrevistas Cognitivas com Experimentos Aleatorizados;
4. Tema Banheiro e Sanitário, realizado por ocasião do Censo Demográfico 2010, para avaliar os impactos e a pertinência das alterações propostas para as perguntas, face à influência de termos regionais sobre a interpretação do

entrevistado. O método utilizado foi o teste de confiabilidade e validade;

5. Tema População Indígena, realizado por ocasião do Censo Demográfico 2010, para verificar a adequação do conjunto de perguntas a serem aplicadas à realidade das comunidades indígenas e avaliar informações de saúde, educação e trabalho na comunidade indígena. O método utilizado foi o teste de usabilidade.
6. Tema Cor ou Raça, como preparação para o Censo Demográfico 2020, para avaliar a qualidade das informações sobre o tema a partir da percepção do informante em relação a sua cor ou raça e do entendimento do informante acerca das categorias de resposta “branca”, “preta”, “amarela”, “parda” e “indígena”.

Fora do escopo dos levantamentos censitários, foram realizados os seguintes testes:

1. Teste de Avaliação da Formação dos Preços Praticados nos Insumos do SINAPI (2013), cujo objetivo foi avaliar a qualidade dos preços dos insumos coletados através da verificação da qualidade das perguntas, com base no entendimento do informante e de sua capacidade de responder de forma precisa. Para tal, foi feita uma adaptação da metodologia dos testes cognitivos à metodologia de coleta de preços dos insumos do SINAPI.
2. Investigação da diferença entre a taxa de desocupação calculada com base nos dados da Pesquisa Nacional por Amostra de Domicílios Contínua e da Pesquisa Mensal de Emprego (2015). Os objetivos do teste foram avaliar a influência tanto do fluxo do questionário quanto do entendimento das perguntas pelos respondentes nas taxas de desocupação. Os métodos utilizados foram Entrevistas Cognitivas e teste de confiabilidade e validade, através da técnica das formas equivalentes.
3. Testes de usabilidade dos dispositivos móveis de coleta (2014), cujo objetivo foi analisar a facilidade de uso do aplicativo de coleta de dados desenvolvido para dispositivos móveis, a fim de gerar recomendações de projeto para fazer a sua interface mais eficiente e adequada para os entrevistadores.
4. Teste Cognitivo do Módulo sobre Trabalho Forçado na Pesquisa Nacional de Saúde, coordenado pelo consultor David Glejberman, da Organização Internacional do Trabalho (2018), utilizando o método de Entrevistas Cognitivas.

Considerações finais

Nos capítulos anteriores, foram apresentados diversos métodos de testes de questionário para avaliar se as perguntas estão captando a informação desejada com qualidade. Para tal, esses métodos visam assegurar tanto que a formulação e compreensão das perguntas reflita as intenções do pesquisador quanto os respondentes estejam dispostos e aptos a responder de forma consistente, sem haver uma carga excessiva sobre os mesmos

(GROVES et al., 2009).

Conforme descrito no capítulo sobre construção de questionário, após a especificação dos conceitos que se pretende medir, deve-se proceder à elaboração das perguntas. Durante essa fase, recomenda-se que sejam aplicados testes pré-campo de acordo com um ou mais métodos escolhidos, de forma a tornar iterativo o processo de avaliação da qualidade das perguntas. Essa avaliação resultante dos pré-testes deve gerar revisões na formulação das perguntas até que as mesmas estejam em conformidade com os conceitos especificados. Dessa forma, o processo de revisão pode ser entendido como a parte final do processo de elaboração do questionário (OFFICE OF MANAGEMENT AND BUDGET, 2016).

Segundo o EUROSTAT (2006), é vital para o processo de desenvolvimento de uma pesquisa que a estratégia de testes de questionário cubra todas as etapas descritas no presente texto. Porém, a definição de uma estratégia de testes apropriada e eficiente requer certa experiência tanto em testes quanto em princípios de qualidade na produção estatística. Cabe mencionar ainda que tais procedimentos requerem tempo e recursos e, apesar da importância dos testes de questionário para se obter resultados com maior qualidade, as áreas responsáveis pelas pesquisas costumam negligenciar essa ferramenta de melhoria contínua. Portanto, recomenda-se que sejam previstos diversos tipos de testes de questionário tanto no planejamento de novas pesquisas e de novos módulos temáticos em pesquisas correntes, quanto na revisão ou análise de resultados de pesquisas realizadas.

Referências

- AAKER, D.A. et al. Marketing Research, 7th Edition. John Wiley & Sons, Inc. New York, 2001.
- AGNER, L., TAVARES, P. Z. Coleta de Dados Domiciliares com Dispositivos Móveis: O Desafio da Usabilidade. Apresentação no 3º Seminário de Metodologia do IBGE – SMI2014. Rio de Janeiro, 2014.
- BORGES, A.; BARBOSA, A.; MICELI, A. P. M. R.; PEREIRA, N. O. M.; LEMES, R. A.; FERREIRA, R. F. Captando Informações da População Indígena nos Censos Demográficos: o Caso da Aldeia de Santa Rosa do Ocoy. Trabalho apresentado no XVII Encontro Nacional de Estudos Populacionais, ABEP, realizado em Caxambu - MG – Brasil, de 20 a 24 de setembro de 2010.
- BARBOSA, A.; BORGES, A.; DINIZ, A. S.; MICELI, A. P. M. R.; LEMES, R. A.; FERREIRA, R. F. A Contribuição dos Testes e Provas Piloto Conjuntas para o Censo 2010. Trabalho apresentado no XVII Encontro Nacional de Estudos Populacionais, ABEP, realizado em Caxambu - MG – Brasil, de 20 a 24 de setembro de 2010.
- CAPE, P. PHILLIPS, K. Questionnaire length and fatigue effects: the latest thinking and practical solutions. 2015. Disponível em <<http://www.surveysampling.com/site/assets/files/1586/questionnaire-length-and-fatigue-effects-the-latest-thinking-and-racticalsolutions.pdf>>. Acesso em 22 de maio de 2018.
- CRAWFORD, I. M. Food and Agriculture Organization of the United Nations. FAO Regional Office for Africa. Rome, 1997.
- DataONE Education Module: Data Quality Control and Assurance. 2016. Disponível em <http://www.dataone.org/sites/all/documents/L05_DataQualityControlAssurance.pptx>. Acesso em 03 de maio de 2018.
- EUROSTAT. Handbook of recommended practices for questionnaire development and testing in the European Statistical System. 2006.
- EUROSTAT. European Health Interview Survey (EHIS wave 3), Methodological manual. 2018.
- FOWLER, F. J. The Case for More Split-Sample Experiments in Developing Survey Instruments. In Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T, Martin, E., Martin, J., and Singer, E. (eds.), Methods for Testing and Evaluating Survey Questionnaires (pp. 173–188). New York: John Wiley and Sons. 2004.
- GLEJBERMAN, D. Seminário: Testes Cognitivos para o Desenho de Questionários – Objetivos, usos e experiência da OIT. Rio de Janeiro. 2018.
- GLEJBERMAN, D. Workshop de Formação para o Módulo de Trabalho Forçado – Testes Cognitivos: Aspectos Conceituais. Rio de Janeiro, 2018.
- GROVES, R.M., FOWLER, F.J. Jr, COUPER, M.P., LEPKOWSKI, J.M., SINGER, E. and TOURANGEAU, R. Survey Methodology. Hoboken, New Jersey: John Wiley & Sons. 2009.
- IBGE. Código de Boas Práticas das Estatísticas Oficiais. Rio de Janeiro, 2013.
- IBGE. Relatório do Teste de Avaliação da Formação dos Preços Praticados nos Insumos do SINAPI. 2013.
- IBGE. Relatório do Teste Cognitivo e Comparativo PME e PNAD Contínua. 2015.

- IPSOS MORI. Cognitive Testing. Disponível em <<https://www.ipsos-mori.com/contactus/offices/scotland/Research-Techniques/Cognitive.aspx>>. Acesso em 10 de julho de 2015.
- KELSEY, J.L. et al. *Methods in Observational Epidemiology*. Oxford University Press, New York, Oxford, 1996.
- MARCONI. M. A.; LAKATOS, E. M. *Técnicas de pesquisa*. São Paulo: Atlas, 1999.
- MARTINS, G. A. Sobre confiabilidade e validade. *Revista Brasileira de Gestão de Negócios*, v. 8, n. 20, p. 1-12, 2006.
- MOORE, J.C. Self/Proxy Response Status and Survey Response Quality: A Review of Literature. *Journal of Official Statistics*, Vol. 4. No. 2, pp 155-172. 1988.
- MOORE, J.; PASCALE, J.; DOYLE, P.; CHAN, A.; KLEIN GRIFFITHS, J. Using Field Experiments to Improve Instrument Design: The SIPP Methods Panel Project. In Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T, Martin, E., Martin, J., and Singer, E. (eds.), *Methods for Testing and Evaluating Survey Questionnaires* (pp. 189–208). New York: John Wiley and Sons. 2004.
- OFFICE OF MANAGEMENT AND BUDGET. *Statistical Policy Working Paper 47: Evaluating Survey Questions: An Inventory of Methods*. Washington, DC. 2016.
- OLSON, K. An Examination of Questionnaire Evaluation by Expert Reviewers, *Field Methods* 22, 295-318. 2010.
- PRESSER, S.; COUPER, M.P.; LESSLER, J.T.; MARTIN, E.; MARTIN, J.; ROTHGEB, J.; SINGER, E. *Methods for Testing and Evaluating Survey Questions*. In *Methods for Testing and Evaluating Survey Questionnaires*, S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, and E. Singer (eds). New York: John Wiley, 1–22. 2004.
- PRESSER, S.; COUPER, M. P.; LESSLER, J. T.; MARTIN, E.; MARTIN, J.; ROTHGEB, J. M.; SINGER, E. *Methods for Testing and Evaluating Survey Questions*. *Public Opinion Quarterly*, 68(1), 109-130. 2004. Disponível em <http://dx.doi.org/10.1093/poq/nfh008>. Acesso em 25 de maio de 2018.
- REVILLA, M; OCHOA, C. Ideal and Maximum Length for a Web Survey. *International Journal of Market Research*, Vol 59, Issue 5, pp. 557 – 565. 2017. Disponível em <<http://journals.sagepub.com/doi/abs/10.2501/IJMR-2017-039#articleCitationDownloadContainer>>. Acesso em 22 de maio de 2018.
- SELLTIZ, C.; WRIGHTSMAN, L. S.; COOK, S. W. *Métodos de pesquisa nas relações sociais*. São Paulo: Editora Pedagógica Universitária (EPU). 1987.
- SUDMAN, S.; BRADBURN, N. M. *Asking Questions*, pp. 208 - 28. 1973.
- TOURANGEAU, R. Experimental Design Considerations for Testing and Evaluating Questionnaires. In Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, E., Martin, J. and Singer, E., *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken, NJ: John Wiley & Sons. 2004.
- TUCKER, C.; MILLER, L. *The Intrahousehold Communications Study: Family Cohesion and the Level of Knowledge*. ASA. 1993.

UNITED NATIONS. Standards and Guidelines for Statistical Surveys. Office of Management and Budget. September 2006. Disponível em <https://unstats.un.org/unsd/dnss/docs-nqaf/USA_standards_stat_surveys.pdf>. Acesso em 11 de abril de 2018.

UNITED NATIONS. Principles and Recommendations for Population and Housing Censuses: the 2020 Round, Revision 3. New York 2015.

ÜSTUN, T. B; CHATTERJI, S.; MECHBAL, A.; MURRAY, C. J. L. Chapter X - Quality assurance in surveys: standards, guidelines and procedures. Household Sample Surveys in Developing and Transition Countries. 2005.

WILLIS, G.B., SCHECHTER, S., WHITAKER, K. A comparison of cognitive interviewing, expert review, and behavior coding. What do they tell us? Proceedings of the Section on Survey Research methods, American Statistical Association, 28–37, 1999.

ZUKERBERG, A.L., VON THURN, J.C.M. Practical considerations in sample selection for behaviour coding pretests. Proceedings of the Section on Survey Research Methods, Alexandria, VA: American Statistical Association, pp. 1116-1121. 1995. Disponível em <<http://www.census.gov/srd/papers/pdf/az9501.pdf>>. Acesso em 25 de maio de 2018.

Se o assunto é **Brasil**,
procure o **IBGE**.



/ibgecomunica



/ibgeoficial



/ibgeoficial



/ibgeoficial

www.ibge.gov.br 0800 721 8181

