Planejamento e avaliação de modelos de análise fatorial confirmatória

Um tutorial reprodutível com R e Quarto

Pablo Rogers and Sérgio Murilo Georgeto

10 de novembro de 2026

Resumo

Esse artigo-tutorial integra orientações metodológicas para uma Análise Fatorial Confirmatória (AFC) a um projeto executável e reprodutível. O objetivo foi sistematizar decisões analíticas fundamentadas na literatura, demonstrar sua implementação em uma aplicação empírica e disponibilizar os materiais que viabilizem examinar, executar e adaptar o percurso apresentado. A aplicação empírica utiliza a escala WHOQOL-Bref e contempla a seleção do modelo de medida, a análise de poder, o pré-processamento, a estimação robusta com tratamento contínuo dos itens, a avaliação do ajuste e a comparação e modificação de modelos alternativos. A proposta didática vincula cada decisão analítica aos procedimentos computacionais responsáveis pelas tabelas, figuras e estatísticas do artigo. No formato HTML: https://phdpablo.github.io/cfa-tutorial/, blocos expansíveis detalham a finalidade, o procedimento e a execução de cada análise, com códigos comentados e hiperlinks para a documentação das funções. A versão em PDF, por sua vez, preserva a leitura contínua tradicional do artigo científico. Dados, funções reutilizáveis, resultados e documentos-fonte são organizados segundo o modelo ARTE (Article Reproducibility Template & Environment), articulando a implementação em R à redação dinâmica em Quarto. Dessa forma, o conjunto fornece um roteiro estruturado para fundamentar as decisões de uma AFC, reproduzir sua aplicação e reutilizar a arquitetura do projeto na elaboração de outros artigos-tutoriais: https://github.com/phdpablo/cfa-tutorial.

Palavras-chave: Análise Fatorial Confirmatória, Modelagem de Equações Estruturais, Validade da Estrutura Interna, Ciência Aberta, Reprodutibilidade, Quarto, *lavaan*.

1 Introdução

A Análise Fatorial Confirmatória (AFC) constitui uma das principais ferramentas para reunir evidências de validade baseadas na estrutura interna de instrumentos de mensuração (Bandalos, 2018; Hughes, 2018; Sireci & Sukin, 2013). A validade representa a propriedade psicométrica central de um modelo de medida (Furr, 2021), pois responde à questão fundamental da mensuração: o que os instrumentos empíricos estão de fato medindo (Bandalos, 2018). Essa perspectiva remonta à definição clássica de validade, compreendida como o grau em que um teste afere o construto que pretende mensurar

(Bandalos, 2018; Furr, 2021; Sireci & Sukin, 2013; Urbina, 2014), preceito alinhado à visão tripartite tradicional e ainda amplamente empregado na literatura (Widodo, 2018).

A perspectiva tripartite estrutura a avaliação psicométrica a partir de três fontes clássicas de evidência: conteúdo, critério e construto (Bandalos, 2018). A validade de conteúdo examina a representatividade e a adequação dos itens frente ao domínio pretendido (Cohen et al., 2022). A validade de critério analisa a associação entre os escores do teste e critérios externos relevantes, como desempenhos correlatos ou desfechos futuros (Cohen et al., 2022). Por sua vez, a validade de construto investiga em que medida o instrumento capta a variável latente teórica, confrontando hipóteses conceituais com a evidência empírica observada (Cohen et al., 2022).

Formalizada pela *American Psychological Association* (APA) nos *Standards for Educational and Psychological Testing* de 1966 (Bandalos, 2018), essa estrutura tripartite orientou a prática psicométrica por décadas. No entanto, a fragmentação da validade em categorias estanques passou a ser criticada (Sireci & Sukin, 2013), conduzindo a uma reformulação teórica em que as subdivisões clássicas foram integradas a uma concepção unificada (Widodo, 2018).

A publicação dos *Standards* de 1999 (AERA et al., 1999) e sua revisão em 2014 (AERA et al., 2014) consolidaram uma mudança paradigmática: a validade deixou de ser considerada uma propriedade intrínseca do instrumento e passou a concernir à legitimidade da interpretação e do uso pretendido de seus escores (Furr, 2021). Sob essa ótica contemporânea, a validade constitui um conceito unitário fundamentado em múltiplas fontes de evidência empírica e teórica, mobilizadas para respaldar as inferências pretendidas para finalidades específicas (Furr, 2021; Urbina, 2014).

Assim, de acordo com as diretrizes vigentes (AERA et al., 2014), a validade expressa o grau em que as evidências empíricas e os fundamentos teóricos sustentam a interpretação dos escores para os fins concebidos. Trata-se de um julgamento avaliativo integrado acerca da adequação das inferências e das decisões fundamentadas nos resultados das avaliações (Bandalos, 2018; Furr, 2021; Urbina, 2014).

Os *Standards* (AERA et al., 2014) sistematizam cinco fontes de evidência de validade: conteúdo, processos de resposta, relações com variáveis externas, consequências do teste e estrutura interna. As evidências de conteúdo atestam a representatividade dos itens em relação ao domínio examinado (Furr, 2021). Os processos de resposta verificam a aderência entre os mecanismos cognitivos dos respondentes e o construto teórico (Sireci & Sukin, 2013). As relações com variáveis externas avaliam associações convergentes e discriminantes com outros construtos de interesse (Furr, 2021). As consequências do teste analisam os desdobramentos éticos e práticos decorrentes da utilização dos instrumentos (Bandalos, 2018).

As evidências baseadas na estrutura interna dizem respeito à conformidade entre as inter-relações dos itens observados e a arquitetura teórica do construto latente (AERA et al., 2014; Rios & Wells, 2014). Essa dimensão analítica desdobra-se em três eixos fundamentais (Rios & Wells, 2014): dimensionalidade, confiabilidade e invariância de medida.

A dimensionalidade afere se a estrutura fatorial hipotetizada sustenta as inferências extraídas das pontuações (Rios & Wells, 2014). A confiabilidade quantifica a precisão dos escores, tradicionalmente mensurada por indicadores de consistência interna (Rios & Wells, 2014). Já a invariância de medida assegura a equivalência psicométrica dos

parâmetros entre subgrupos populacionais, como sexo ou faixa etária, viabilizando comparações entre os grupos.

Por meio da AFC, examinam-se simultaneamente essas três fontes de evidência da estrutura interna (Bandalos, 2018; Flora & Flake, 2017; Hughes, 2018; Reeves & Marbach-Ad, 2016; Rios & Wells, 2014). Em Ciências Sociais Aplicadas, o pesquisador usualmente dispõe de hipóteses teóricas prévias sobre a dimensionalidade do construto (Sireci & Sukin, 2013), contexto no qual a AFC é recomendada para confrontar o modelo de medida teórico com a matriz de covariâncias empírica (Rios & Wells, 2014).

A AFC constitui o modelo de mensuração da modelagem de equações estruturais baseada em covariâncias (CB-SEM, do inglês *Covariance-Based Structural Equation Modeling*) (Brown, 2023; Harrington, 2009; Jackson et al., 2009; Kline, 2023; Nye, 2022). A SEM integra procedimentos de regressão múltipla e análise fatorial em uma única estrutura analítica (Hoyle, 2023; Kline, 2023), permitindo estimar relações complexas entre construtos teóricos enquanto modela expressamente o erro de mensuração (Hoyle, 2023).

A condução da AFC envolve escolhas metodológicas condicionadas pela natureza dos dados, pelo formato de resposta dos itens e pelos objetivos do estudo (Crede & Harms, 2019; Flake et al., 2017; Flake & Fried, 2020; Jackson et al., 2009; Nye, 2022; Rogers, 2024). Tais decisões impactam diretamente a viabilidade analítica: a complexidade do modelo, a distribuição das categorias de resposta e o tamanho amostral condicionam a precisão das estimativas e a convergência dos algoritmos de estimação (Rhemtulla et al., 2012). Dessa forma, ponderar as implicações psicométricas dos instrumentos ainda no desenho da pesquisa previne inconsistências e dificuldades operacionais nas fases posteriores.

Nesse contexto, formula-se a seguinte questão de pesquisa: **quais escolhas metodológicas tornam a aplicação da AFC mais viável e flexível para obter evidências de validade baseadas na estrutura interna?** Cabe enfatizar que as decisões técnicas atuam como instrumentos subordinados aos objetivos teóricos da investigação (Pilcher & Cortazzi, 2023). São os objetivos teóricos que delimitam o problema de pesquisa e direcionam a seleção metodológica, cabendo ao método viabilizar o teste rigoroso das hipóteses formuladas.

Na Psicometria, diversas questões metodológicas admitem mais de uma abordagem defensável. Nesses casos, a escolha deve considerar os objetivos da pesquisa e as consequências da modelagem para a interpretação dos resultados. Nesse artigo, adotamos uma recomendação prática: entre instrumentos adequados ao construto e à população de interesse, privilegiam-se aqueles cujos itens possam ser tratados como contínuos. Essa decisão deve ser planejada e fundamentada, pois oferece acesso às propriedades analíticas da máxima verossimilhança (ML) sem impor o formato ordinal como um obstáculo metodológico intransponível (Rhemtulla et al., 2012; Robitzsch, 2020, 2022).

Para esse propósito, recomendam-se instrumentos com itens estruturados em cinco ou mais categorias de resposta, preferencialmente sete, desde que disponham de evidências prévias de validade para o uso pretendido. Estudos de simulação em AFC sustentam o emprego da ML robusta com cinco a sete categorias em diversas condições empíricas, com resultados particularmente favoráveis quando os limiares das categorias são aproximadamente simétricos (Rhemtulla et al., 2012).

O tratamento contínuo permite empregar a estimação por ML e suas correções robustas,

além de articular o modelo a rotinas flexíveis de avaliação de ajuste, comparação entre estruturas concorrentes e tratamento direto de dados ausentes (Rhemtulla et al., 2012; Robitzsch, 2020). A modelagem ordinal, por sua vez, também envolve pressupostos distributivos: a abordagem convencional baseada em matrizes de correlação policórica assume variáveis latentes contínuas subjacentes com distribuição normal bivariada. Assim, ambas as abordagens operam sob premissas teóricas específicas, cabendo ao pesquisador justificar a adequação da representação matemática escolhida frente aos dados disponíveis e às metas analíticas (Robitzsch, 2020, 2022).

Essas propriedades tornam o tratamento contínuo uma alternativa consistente para o presente artigo-tutorial. Na presença de dados omissos, por exemplo, a máxima verossimilhança com informação completa (*Full Information Maximum Likelihood*, FIML) viabiliza sua incorporação direta na estimação, atendidas as condições teóricas sobre o modelo e o mecanismo de ausência. A literatura indica, adicionalmente, cenários nos quais a ML robusta exibe desempenho superior em amostras menores quando comparada aos estimadores de mínimos quadrados ponderados (Rhemtulla et al., 2012).

O presente artigo-tutorial tem como objetivo sistematizar orientações metodológicas para a condução da AFC, demonstrar sua implementação computacional em uma aplicação empírica e disponibilizar um projeto executável e reprodutível que integre dados, rotinas analíticas e redação dinâmica dos resultados. O percurso abrange a seleção do modelo de medida, a análise de poder (AP), o pré-processamento dos dados, a estimação dos parâmetros, a avaliação do ajuste global e local e a comparação e modificação dos modelos alternativos. Cada decisão metodológica é discutida no referencial teórico sob uma tríplice perspectiva: o que fazer (Flake et al., 2017; Nye, 2022; Rogers, 2024), o que evitar (Crede & Harms, 2019; Rogers, 2024) e o que reportar (Flake & Fried, 2020; Jackson et al., 2009; Rogers, 2024).

A aplicação empírica utiliza a escala WHOQOL-Bref e retoma a base de dados analisada por Rogers (2024) e Rogers (2022), documentando os procedimentos desde os registros brutos. A implementação computacional em R emprega o ecossistema do *lavaan* (Rosseel, 2023) e rotinas modulares reutilizáveis para estruturar operações repetitivas. Tabelas, figuras e estatísticas são compiladas pelo código associado ao manuscrito em Quarto, permitindo ao leitor verificar como cada decisão teórica se traduz nos resultados reportados.

O público-alvo compreende pesquisadores e estudantes de pós-graduação com noções fundamentais de estatística multivariada e CB-SEM, contemplando diferentes níveis de proficiência em programação. Para atender a essa audiência, o tutorial articula as justificativas teóricas aos procedimentos estatísticos, explicitando os códigos junto às saídas correspondentes. Essa estrutura facilita a compreensão do processo e sua execução prática, franqueando também aos leitores experientes o acesso a materiais modulares para adaptação em suas próprias pesquisas.

A função essencial de artigos-tutoriais reside em aproximar métodos estatísticos e computacionais avançados de pesquisadores não especialistas (Martins, 2021; Mendes-Da-Silva, 2023). Nessa proposta, a contribuição didática estrutura-se em três frentes: propiciar o entendimento crítico das decisões de uma AFC; assegurar a reprodutibilidade integral da aplicação empírica; e oferecer uma organização de projeto reutilizável na elaboração de novos tutoriais. Protocolos transparentes, dados e códigos abertos viabilizam o escrutínio das escolhas realizadas e suprem o ensino de métodos quantitativos de recursos aplicados, em estrita consonância com os preceitos da Ciência Aberta (Gilroy & Kaplan, 2019;

Kathawalla et al., 2021; Klein et al., 2018) e com os princípios FAIR (*Findable, Accessible, Interoperable, and Reusable*) (Wilkinson et al., 2016).

A prática da Ciência Aberta pressupõe a incorporação ativa de rotinas de documentação, compartilhamento e governança de dados à atividade científica (Limongi & Rogers, 2025a). Sua consolidação em um fluxo de trabalho exige sistematizar materiais, rastrear versões e conectar as análises estatísticas ao texto final (Limongi & Rogers, 2025b). Para operacionalizar essa integração, adota-se o ARTE (Article Reproducibility Template & Environment), que congrega a gestão do projeto, o controle de versionamento, a administração de dependências computacionais e a editoração de documentos reproduzíveis (Rogers & Limongi, 2025).

O ARTE propõe a adoção progressiva dessas práticas, estendendo-se da organização inicial e disponibilização dos arquivos até o isolamento completo do ambiente computacional por meio de contêineres, em conformidade com a complexidade do estudo e a familiaridade do pesquisador com ferramentas de desenvolvimento (Rogers & Limongi, 2025). No tutorial, essa estrutura conecta os dados brutos aos procedimentos de tratamento e modelagem, aos resultados consolidados e ao texto do manuscrito. Os materiais compartilhados possibilitam reproduzir passo a passo o percurso metodológico e transpor essa infraestrutura para outros problemas psicométricos.

O artigo é disponibilizado em dois formatos complementares. O PDF preserva a narrativa científica linear, contendo as fundamentações teóricas, os resultados empíricos e suas interpretações dos resultados. O HTML acrescenta uma camada didática interativa: abaixo de tabelas e gráficos, blocos expansíveis detalham a finalidade, a preparação, o procedimento estatístico e a execução das rotinas. O conteúdo desses blocos elucida como os scripts geram cada saída numérica, incorporando comentários objetivos nas linhas de código e hiperlinks para a documentação oficial das funções utilizadas. As rotinas autorais encontram-se documentadas nos scripts do projeto, permitindo modular a profundidade da leitura conforme o interesse do leitor.

Além dessa introdução, o artigo compreende uma fundamentação teórica, seis etapas aplicadas e a seção de conclusão. A Seção 2 reúne as recomendações metodológicas que balizam a aplicação empírica. A Seção 3 apresenta o instrumento psicométrico e as especificações dos modelos alternativos; a Seção 4 aborda o planejamento amostral por meio da AP; e a Seção 5 detalha a obtenção da amostra analítica e a análise descritiva dos indicadores. A estimação dos parâmetros e a avaliação da convergência das soluções são examinadas na Seção 6. A Seção 7 congrega evidências de ajuste global, confiabilidade e diagnóstico local de resíduos, subsidiando os procedimentos comparativos e as reespecificações da Seção 8. Por fim, a Seção 9 sintetiza as principais considerações do trabalho. Essa organização adota as diretrizes para artigos-tutoriais propostas por Martins (2021), integrando os preceitos metodológicos a rotinas abertas que viabilizam a replicação autônoma do estudo.

2 Fundamentação das decisões analíticas

O presente tutorial retoma o contexto de aplicações discutido por Rogers (2024): modelos de fator comum, com estrutura multifatorial definida e propriedades psicométricas previamente investigadas, frequentemente empregados como etapa de uma pesquisa mais ampla. A diferença reside na estratégia analítica aqui recomendada: planejar a escolha de instrumentos cujos itens permitam o tratamento contínuo, utilizando a máxima

verossimilhança robusta (MLR) como método de estimação. A fundamentação dessa escolha exige distinguir o formato de resposta, o objeto da análise e os pressupostos do modelo.

Itens individuais e escores agregados de uma escala não são intercambiáveis nessa discussão. A possibilidade de analisar uma soma ou média de itens como contínua não basta para justificar o mesmo tratamento de cada indicador em uma AFC. Harpe (2015) estabelece essa distinção e demonstra que itens individuais com formatos numéricos de pelo menos cinco categorias podem, em geral, ser tratados como contínuos. Norman (2010) complementa essa perspectiva ao evidenciar que a avaliação de procedimentos paramétricos deve considerar sua robustez, em vez de rejeitá-los automaticamente pela classificação ordinal dos dados. Essas contribuições respaldam a discussão geral. A justificativa específica para a AFC decorre dos estudos sobre seus modelos e estimadores.

Nessa direção, Rhemtulla et al. (2012) identificou desempenho satisfatório da ML robusta com cinco a sete categorias em diversas condições, embora a assimetria dos limiares e a magnitude das cargas fatoriais influenciassem os resultados. Com sete categorias e limiares aproximadamente simétricos, a abordagem contínua frequentemente apresentou desempenho equivalente ou superior ao da alternativa categórica examinada. Esses resultados fornecem suporte à diretriz prática adotada no presente artigo.

Sob outra perspectiva, Robitzsch (2020) e Robitzsch (2022) argumentam que a escolha entre modelar as relações entre os escores observados e modelar respostas contínuas subjacentes envolve diferentes definições dos parâmetros de interesse. A modelagem ordinal não é universalmente superior apenas porque os itens apresentam categorias ordenadas. Assim, a recomendação adotada nesse tutorial combina a possibilidade de defender o modelo contínuo com evidências de desempenho e vantagens operacionais. Propõe-se esse caminho desde a fase de planejamento, mantendo a adequação do instrumento ao objetivo da pesquisa como critério prioritário.

As seções seguintes aprofundam aspectos metodológicos discutidos na literatura (Crede & Harms, 2019; Flake et al., 2017; Flake & Fried, 2020; Jackson et al., 2009; Nye, 2022; Rogers, 2024), sintetizados por Rogers (2024) no contexto de modelos ordinais. Esses tópicos são aqui estruturados sob a perspectiva do que fazer (Flake et al., 2017; Nye, 2022; Rogers, 2024), do que evitar (Crede & Harms, 2019; Rogers, 2024) e do que reportar (Flake & Fried, 2020; Jackson et al., 2009; Rogers, 2024), com foco em modelos de medida cujos indicadores admitem tratamento contínuo.

Assume-se que o leitor possui conhecimento prévio sobre o assunto, remetendo-se detalhes teóricos introdutórios aos manuais clássicos (Brown, 2015; Kline, 2023) ou a artigos que oferecem introdução à CB-SEM (Davvetas et al., 2020; Shek & Yu, 2014). O contexto analítico circunscreve-se à AFC no âmbito da CB-SEM (Brown, 2015; Jackson et al., 2009; Kline, 2023; Nye, 2022). Não se abordam especificações na SEM baseada em variâncias (VB-SEM), encontradas principalmente na literatura sobre *Partial Least Squares* SEM (PLS-SEM) (J. F. Hair et al., 2017, 2022; Henseler, 2021).

2.1 Seleção do Modelo de Medida

Na seleção do modelo de medida, recomenda-se privilegiar instrumentos cujos itens tenham cinco ou mais categorias de resposta, preferencialmente sete, entre as alternativas com evidências de validade para o construto, a língua e a população de interesse. A preferência por sete categorias decorre do desempenho favorável observado em condições

simuladas de AFC e da possibilidade de maior diferenciação das respostas, sem representar uma garantia de maior poder estatístico ou de mensuração ótima em qualquer contexto (Harpe, 2015; Rhemtulla et al., 2012). Essa orientação aplica-se à escolha entre instrumentos apropriados e não autoriza alterar arbitrariamente o formato de resposta de uma escala validada, pois modificações exigem nova avaliação de suas propriedades psicométricas (Harpe, 2015).

O número de categorias oferecidas também deve ser confrontado com a forma como são utilizadas pelos respondentes. Um item com cinco ou sete opções pode apresentar respostas concentradas em apenas duas ou três categorias, como destaca Harpe (2015). Assim, recomenda-se examinar, no planejamento, evidências de aplicações anteriores na população-alvo e, após a coleta, as frequências das categorias e a distribuição de cada item. A descrição dos indicadores no pré-processamento documenta essas características e subsidia a avaliação da estratégia de estimação. As correções robustas não dispensam essa análise nem corrigem automaticamente uma especificação inadequada do modelo de medida (Rhemtulla et al., 2012).

Ao selecionar um modelo de medida, convém priorizar escalas já validadas na língua de aplicação e testadas na população-alvo da pesquisa (Flake et al., 2017). Adicionalmente, cumpre verificar se essas escalas não são proprietárias ou privativas de determinada categoria profissional. Recomenda-se consultar sistemas oficiais de avaliação de testes psicológicos ou recorrer a instrumentos desenvolvidos por órgãos públicos, consórcios acadêmicos e centros de pesquisa, que comumente contam com respaldo metodológico e investimento substancial em validação para políticas públicas.

A prospecção do modelo de medida deve abranger periódicos especializados, manuais técnicos e teses de referência na área. Schumacker et al. (2021) fornecem uma listagem sistemática para busca de instrumentos como ponto de partida. Existem diversas revisões sistemáticas e meta-análises da literatura sobre modelos de medida voltados a construtos específicos. É indispensável examinar o estudo seminal da escala e suas validações empíricas subsequentes. Kline (2016, p. 29) apresenta um roteiro sistemático para avaliação de modelos de medida potenciais.

Devem-se evitar instrumentos elaborados *ad hoc* sem rigor psicométrico formal (Flake & Fried, 2020; Kline, 2016), visto que a mera existência de uma escala não assegura sua validade (Flake et al., 2017). Tampouco se recomenda adotar instrumentos pouco referenciados ou traduções diretas desprovidas de adaptação transcultural rigorosa em amostras independentes. Do mesmo modo, não se devem ocultar estruturas fatoriais concorrentes documentadas na literatura (como modelos hierárquicos ou bifator), pois a consideração de representações alternativas respaldadas teoricamente qualifica o escopo confirmatório da investigação (Crede & Harms, 2019).

Cabe fundamentar as propriedades psicométricas da escala selecionada, indicando evidências empíricas prévias de aplicação na população almejada, bem como coeficientes de confiabilidade e validade (Flake & Fried, 2020; Jackson et al., 2009; Kline, 2016). Se para uma mesma escala existirem diferentes estruturas fatoriais esperadas, devem-se justificar as comparações que serão empreendidas para assegurar o caráter genuinamente confirmatório da AFC (Jackson et al., 2009).

Recomenda-se especificar o modelo principal e apoiar a escolha discutindo seus méritos em relação a modelos teóricos equivalentes e concorrentes (Jackson et al., 2009), com representação gráfica das estruturas testadas (Jackson et al., 2009). Definem-se *a*

priori, com base na literatura empírica prévia, eventuais cargas cruzadas (*cross-loadings*) teoricamente plausíveis (Brown, 2023; Nye, 2022).

Considerando que a escala selecionada frequentemente integra um questionário de pesquisa mais amplo, convém adotar as recomendações procedimentais para triagem de respostas descuidadas ou com esforço insuficiente (*Careless or Insufficient Effort Responding*, C/IER) (Curran, 2016; Huang et al., 2026; Ward & Meade, 2023). Recomenda-se planejar desde o desenho do instrumento a inclusão de itens de atenção explícita (*instructed response items* e, quando apropriado, *bogus/infrequency items*) e o registro do tempo de resposta do questionário, preferencialmente por página (*page time*). Esse protocolo possibilita identificar respondentes que não leram os itens ou que completaram o questionário em velocidade incompatível com um processamento cognitivo atento (Ward & Meade, 2023).

2.2 Análise de Poder

No âmbito da AP, a primeira diretriz metodológica consiste em rejeitar decisões de tamanho amostral fundamentadas em regras de bolso heurísticas. Na literatura, encontram-se recomendações empíricas díspares, que variam de tamanhos mínimos absolutos (N de 50 a 1.000) a razões arbitrárias entre observações e parâmetros livres (N/q de 5:1 a 20:1) (Kline, 2023; T. A. Kyriazos, 2018). Tais parâmetros baseiam-se em cenários de simulação particulares e dificilmente refletem as especificidades de um modelo empírico concreto.

Tampouco se deve recorrer a estimativas de limite inferior como substituto da AP voltada a efeitos específicos (Westland, 2010). O tamanho de amostra mínimo genérico tende a resultar em poder estatístico insuficiente para detectar efeitos-alvo no modelo sob investigação (Wang, 2023). Nesse contexto, desaconselha-se o emprego da calculadora de Soper, amplamente difundida em pesquisas aplicadas, cujas estimativas baseadas em fórmulas analíticas genéricas não captam as nuances estruturais da CB-SEM.

Por se tratar de uma técnica multivariada complexa, regras gerais de tamanho amostral em SEM são inadequadas (T. A. Kyriazos, 2018). A prática contemporânea preconiza planejar o poder estatístico de modo customizado para cada estudo, acomodando o desenho da pesquisa e suas hipóteses centrais (Feng & Hancock, 2023; Moshagen & Bader, 2024). Não existe regra universal para determinação do tamanho amostral em modelagem estrutural (Kline, 2023).

Embora seja possível estimar o poder *a posteriori*, o planejamento amostral deve ser conduzido preferencialmente *a priori* (Feng & Hancock, 2023). Esse procedimento induz o pesquisador a aprofundar o domínio conceitual do estudo (Leite et al., 2023), pois exige especificar a magnitude esperada dos parâmetros populacionais com base na literatura subjacente (Feng & Hancock, 2023). A fundamentação teórica prévia simplifica significativamente a condução da AP *a priori*.

A AP *a priori* permite estimar o tamanho amostral necessário para atingir o poder estatístico desejado sob os pressupostos do modelo. Essa abordagem eleva a probabilidade de obter resultados replicáveis e comparáveis com a literatura pregressa, favorecendo futuras meta-análises. Adicionalmente, otimiza o orçamento da pesquisa ao evitar o desperdício de recursos associado a coletas superdimensionadas (Feng & Hancock, 2023).

Em SEM, a AP pode ser realizada por derivação analítica, com base na teoria assintótica, ou por simulação computacional. A abordagem analítica requer definir o efeito de interesse

em termos de um indicador diretamente associado ao parâmetro de não centralidade. Por sua vez, a AP baseada em simulação obtém uma estimativa empírica do poder a partir de um modelo populacional especificado (Moshagen & Bader, 2024). O poder é relevante tanto para a avaliação do ajuste global quanto para a significância de parâmetros individuais, sendo viável aplicar ambas as abordagens a esses dois objetivos (Feng & Hancock, 2023).

Em uma AFC conduzida para validar um modelo de medida que subsidiará etapas posteriores, recomenda-se avaliar no mínimo o poder associado ao teste de ajuste global (Nye, 2022; Wang, 2023). Não obstante as limitações das formulações analíticas (Feng & Hancock, 2023), elas podem servir como ponto de partida exploratório e complemento aos métodos de simulação (Moshagen & Bader, 2024).

Para a implementação analítica da AP do ajuste global, dispõe-se de soluções baseadas no teste de qui-quadrado não central (proposta de Saris e Satorra) e na estatística RMSEA sob modelos correta ou incorretamente especificados (proposta de MacCallum e colaboradores) (Wang, 2023). Tais alternativas encontram-se implementadas em aplicações interativas desenvolvidas por Jak et al. (2021), Moshagen & Bader (2024) e Wang & Rhemtulla (2021), bem como na plataforma WebPower de Zhang & Yuan (2018).

O ambiente WebPower apresenta soluções baseadas em Simulação de Monte Carlo (SMC), contemplando dados omissos e não normais por meio de interface gráfica. A solução de Wang & Rhemtulla (2021) também incorpora rotinas de simulação, e ambas permitem estimar o poder para parâmetros específicos do modelo. Revisões abrangentes desses métodos analíticos encontram-se em Feng & Hancock (2023), Jak et al. (2021), Nye (2022) e Wang (2023).

Para a AP em AFC, a estratégia mais consistente reside em executar uma SMC com ferramentas que assegurem a documentação e a reprodutibilidade dos parâmetros. Mesmo para soluções analíticas exploratórias, recomenda-se o ecossistema R, com os pacotes *semTools* (Jak et al., 2021) e *semPower* (Jobst et al., 2023; Moshagen & Bader, 2024).

Para a condução da SMC em modelagem estrutural, o pacote *simsem* (Pornprasertmanit et al., 2022) oferece uma alternativa flexível. Essa recomendação apoia-se em quatro justificativas metodológicas.

Em primeiro lugar, a abordagem analítica clássica repousa sobre distribuições amostrais conhecidas e propriedades assintóticas dos testes estatísticos. Essas premissas teóricas são facilmente flexibilizadas na SMC, cuja popularidade decorre da capacidade de acomodar múltiplos cenários empíricos (dados categóricos, omissões, estruturas longitudinais e estimadores robustos de máxima verossimilhança) (Feng & Hancock, 2023).

Em segundo lugar, a SMC exige que o pesquisador explicite a matriz populacional de parâmetros do modelo de medida, fortalecendo a validade interna das inferências analíticas (Leite et al., 2023). Os resultados simulados auxiliam, inclusive, a calibrar empiricamente os pontos de corte dos índices de ajuste para as fases posteriores.

Em terceiro lugar, a objeção de que a SMC impõe custo computacional proibitivo em modelos complexos (Nye, 2022) perde sustentação na prática cotidiana de pesquisa (Byrne, 2016). Em aplicações usuais de AFC, os modelos não atingem ordens de complexidade intratáveis, e a capacidade atual de processamento viabiliza soluções escaláveis. A otimização algorítmica e o parcelamento da simulação em etapas reduzem a

sobrecarga computacional (Leite et al., 2023), sendo a SMC a única alternativa confiável para modelos estruturais densos (Moshagen & Bader, 2024).

Por fim, a lógica da SMC é conceitualmente intuitiva e acessível a pesquisadores com formação estatística intermediária (Leite et al., 2023; Nye, 2022; Wang, 2023; Wang & Rhemtulla, 2021). Documentações aplicadas e tutoriais estruturados facilitam seu emprego na SEM, como exemplificado pelo pacote *simsem*, cuja sintaxe se integra nativamente ao padrão do pacote *lavaan*. Para pesquisadores que dominam essa interface, a flexibilidade analítica da simulação compensa amplamente o investimento técnico inicial (Feng & Hancock, 2023).

A documentação transparente do planejamento amostral para o modelo de medida fortalece a reprodutibilidade científica (Flake et al., 2017; Flake et al., 2022), contribuindo para acumular evidências de validade da estrutura interna (Flake & Fried, 2020) e externa (Leite et al., 2023). Na AFC, um poder estatístico inadequado não detectado compromete a validade das inferências do modelo estrutural subsequente. A SMC permite avaliar o poder sob cenários relevantes para o modelo de medida, lembrando que estimativas precisas de parâmetros irrelevantes são menos úteis do que estimativas aproximadas de fenômenos teoricamente relevantes (Leite et al., 2023).

2.3 Pré-processamento

A qualidade dos dados constitui condição necessária, embora não suficiente, para que as estimativas de uma AFC sejam válidas e interpretáveis. Além do controle de inconsistências técnicas (ver Whittaker & Schumacker (2022)) e do viés do método comum (*Common Method Bias*, CMB; ver Podsakoff et al. (2012)), o pré-processamento deve contemplar sistematicamente as respostas descuidadas ou com esforço insuficiente (C/IER). O termo C/IER designa preenchimentos realizados sem atenção adequada ao enunciado dos itens, abrangendo padrões aleatórios e não aleatórios, como marcações invariantes na mesma opção (*straightlining*), respostas concentradas no ponto médio e sequências repetitivas (Curran, 2016; Huang et al., 2015).

Cumpre distinguir o C/IER de outros vieses de resposta: a desejabilidade social e a aquiescência, por exemplo, afetam primariamente a validade baseada nos processos de resposta, conforme discutido na introdução desse trabalho. O C/IER atua em nível estrutural: quando respostas desatentas contaminam múltiplos instrumentos autorrelatados na mesma direção, introduzem variância compartilhada espúria, operando como fonte de *Common Method Variance* (CMV) que distorce correlações bivariadas, cargas fatoriais e inferências estatísticas (Huang et al., 2015). Assim, o C/IER constitui uma causa contingente de CMB, distinguindo-se do erro aleatório simples de mensuração. A revisão de Huang et al. (2026) demonstra que a triagem de C/IER ainda carece de padronização, reforçando a necessidade de sua incorporação protocolar ao fluxo de pré-processamento.

Com os dados primários organizados em formato legível por máquina (como arquivos de texto plano com separação por vírgula ou tabulação), a etapa inicial consiste em expurgar os participantes com desistência prematura do preenchimento. A permanência inadvertida desses registros durante as fases preliminares distorce as análises de dados omissos, a identificação de observações discrepantes e os diagnósticos de colinearidade.

Itens de atenção explícita (*instructed response items* ou *infrequency items*) e o tempo de resposta do questionário, preferencialmente mensurado por página (*page time*), representam procedimentos clássicos de controle que devem ser previstos no desenho do

instrumento (Curran, 2016; Huang et al., 2026). Os critérios de corte temporal devem considerar as peculiaridades do questionário. A literatura estabelece limiares de dois a quatro segundos por item: dois segundos configuram referência para assinalar leituras inviáveis (Curran, 2016), enquanto Jones et al. (2023) avaliam limites de três a quatro segundos por assertiva.

No âmbito dos métodos *post hoc* de detecção de C/IER, calcula-se o índice de variabilidade intraindividual de resposta (*Intraindividual Response Variability*, IRV) para avaliar a dispersão das respostas do participante. O IRV corresponde ao desvio-padrão das respostas de cada indivíduo; valores iguais ou muito próximos de zero assinalam a presença de *straightlining*, padrão sistemático que inflaciona artificialmente a consistência interna da escala. Complementarmente, o índice *Longstring* quantifica a maior sequência ininterrupta de respostas idênticas, exibindo sensibilidade para assinalar indivíduos que escolhem repetidamente a mesma alternativa (Curran, 2016; Hong et al., 2020). Valores atipicamente elevados de IRV sugerem, por sua vez, padrões de resposta aleatórios, incompatíveis com a covariância esperada entre itens de um mesmo construto (Curran, 2016; Hong et al., 2020).

A identificação de observações discrepantes (*outliers*) deve ser integrada ao protocolo de triagem de C/IER. Padrões aberrantes de resposta comprometem a matriz de covariâncias e podem desestabilizar a solução fatorial previamente à estimação do modelo. A distância de Mahalanobis constitui o indicador multivariado padrão: mensura a discrepância entre o vetor de respostas de cada participante e o centroide da amostra, apresentando elevada sensibilidade para identificar respostas aleatórias (Curran, 2016; Hong et al., 2020).

Recomenda-se adotar um subconjunto complementar de indicadores (*best subset approach*), combinando, por exemplo, um critério procedural, o IRV e a distância de Mahalanobis, visto que cada métrica capta facetas distintas de C/IER (Hong et al., 2020). A utilização indiscriminada de todos os índices disponíveis (*kitchen sink approach*) degrada a especificidade do procedimento sem ganhos proporcionais de sensibilidade.

A articulação dos indicadores também pode ser operacionalizada pela abordagem de múltiplos obstáculos (*multiple-hurdle approach*), na qual a infração a qualquer critério pré-especificado determina a remoção do registro (Curran, 2016; DeSimone et al., 2015). Esse método não exige que o indivíduo seja flagrado simultaneamente por todas as métricas. A definição dos limiares deve ser documentada, separando o expurgo de preenchimentos descuidados contra o risco de exclusão indevida de respondentes atentos. Em relação aos itens de atenção, deve-se explicitar quantas falhas configuram desatenção efetiva, evitando classificar erros isolados como descuido deliberado.

Quanto à multicolinearidade, recomenda-se calcular o fator de inflação da variância (VIF) dos indicadores observados e investigar magnitudes atípicas, que evidenciam redundância informacional e potenciais dificuldades de convergência (Whittaker & Schumacker, 2022). Em modelos reflexivos de AFC, a correlação entre itens de um mesmo fator é teoricamente desejável (Grewal et al., 2004), mas não deve atingir patamares que provoquem singularidade na matriz de covariâncias (Flora et al., 2012; Kline, 2023; Whittaker & Schumacker, 2022). A elevada colinearidade constitui causa frequente de matrizes não positivas definidas (Flora et al., 2012; Kline, 2023), podendo acarretar interpretações enviesadas dos parâmetros (Marsh et al., 2004) e sugerir ausência de validade discriminante (Grewal et al., 2004).

O tratamento de dados omissos (*missings*) pode ser operacionalizado na fase de estimação.

A eliminação por listagem (*listwise*) ou por pares (*pairwise*) distingue-se de estratégias de imputação pontual única, como a substituição pela média ou por predição de regressão. A imputação simples desconsidera a incerteza estatística dos valores imputados, subestimando os erros-padrão das estimativas (Enders, 2023). Em contrapartida, a imputação múltipla e a máxima verossimilhança com informação completa (FIML) representam abordagens metodologicamente rigorosas, cuja escolha depende da estrutura do modelo e do mecanismo subjacente de perda dos dados (Kline, 2023).

A abordagem FIML configura a escolha padrão em modelos de AFC estimados por ML ou MLR, pois incorpora diretamente as informações observadas de cada participante na função de verossimilhança, dispensando a geração e agregação de múltiplos bancos imputados (Brown, 2023). Seu emprego pressupõe dados faltantes completamente ao acaso (MCAR) ou ao acaso (MAR), condicionados às variáveis observadas, além de correta especificação estrutural (Enders, 2023; Kline, 2023). A imputação múltipla permanece como alternativa válida, embora demande decisões adicionais sobre o modelo de imputação e a combinação das estimativas (Kline, 2023).

A disponibilidade de procedimentos robustos de estimação reduz a necessidade de aplicar transformações numéricas aos dados com a finalidade exclusiva de forçar aderência à normalidade multivariada (Brown, 2023). Recomenda-se, igualmente, evitar a criação de parcelas de itens (*item parceling*), pois essa agregação mascara a dimensionalidade real da escala e prejudica a detecção de más especificações estruturais (Brown, 2023; Crede & Harms, 2019).

Inspeções descritivas e gráficas dos indicadores constituem prática indispensável na análise preliminar. A literatura detalha diversos procedimentos de diagnóstico descritivo (Brown, 2023; Collier, 2020; Flora et al., 2012; J. Hair et al., 2019; Kline, 2023; Whittaker & Schumacker, 2022), dispensando-se repetições extensivas para os propósitos do presente manuscrito.

Em suma, cabe relatar detalhadamente todos os protocolos de saneamento dos dados. A documentação dos dados omissos deve ser transparente, especificando-se a extensão e a estratégia de tratamento adotada (Crede & Harms, 2019). É imperativo reportar os critérios de triagem de C/IER, com os índices, pontos de corte e proporção de exclusões (Huang et al., 2026), bem como os diagnósticos de multicolinearidade prévios à AFC. Por fim, devem-se informar as versões dos softwares e pacotes estatísticos empregados (Flake & Fried, 2020; Jackson et al., 2009) e disponibilizar os dados brutos ou a matriz de covariâncias dos itens em repositório aberto (Crede & Harms, 2019; Flake et al., 2022; Flake & Fried, 2020).

2.4 Processo de Estimação

Na AFC com itens ordinais (escalas Likert de até cinco pontos), Rogers (2024) argumenta, com respaldo na literatura psicométrica, em favor de estimadores da família dos mínimos quadrados ordinários, preferencialmente o ULS (*Unweighted Least Squares*) robusto (RULS) para amostras reduzidas, ou o DWLS (*Diagonally Weighted Least Squares*) robusto (RDWLS) para amostras ampliadas.

No entanto, conforme discutido na introdução desse trabalho, evidências empíricas (Rhemtulla et al., 2012; Robitzsch, 2022) e fundamentos teóricos (Robitzsch, 2020) indicam que o tratamento contínuo de itens ordinais produz estimativas consistentes quando a escala apresenta cinco ou mais categorias de resposta. Com seis a sete categorias,

os estudos registram convergência entre os métodos e recomendam formatos de resposta mais amplos (≥ 5) para elevar o poder de rejeitar modelos incorretos (Maydeu-Olivares et al., 2017).

Apesar de o estimador de máxima verossimilhança (ML) ser robusto a desvios moderados de normalidade (Brown, 2015), em condições nas quais assimetria e curtose pronunciadas constituem a regra e não a exceção, justifica-se a utilização de variantes com correção robusta, particularmente o MLR (Robitzsch, 2020).

Existem múltiplos algoritmos para obter erros-padrão robustos e estatísticas de teste corrigidas com propriedades assintóticas equivalentes (Kline, 2023). No entanto, o MLR exibe aplicabilidade mais ampla, especialmente na estimação integrada com dados omissos via FIML (Brown, 2015; Rosseel, 2012). Adicionalmente, acumulam-se evidências empíricas que corroboram a consistência e a eficácia do MLR frente a estimadores concorrentes (Bandalos, 2014; Holgado-Tello et al., 2016; Li, 2016a; Nalbantoğlu-Yılmaz, 2019; Yang & Liang, 2013; Yang-Wallentin et al., 2010).

Na fase de diagnóstico da estimação, cumpre inspecionar prioritariamente a ocorrência de casos Heywood (*Heywood cases*), como cargas fatoriais padronizadas superiores à unidade ou variâncias residuais negativas (Nye, 2022). Cabe ao pesquisador justificar a escolha do método de estimação com base na natureza dos dados e na estrutura do modelo (Crede & Harms, 2019), relatando eventuais problemas de convergência ou de identificação paramétrica (Nye, 2022). Diante de não convergência, devem-se examinar a identificação formal, as restrições do modelo e a matriz de entrada antes de cogitar alterações empíricas no estimador. Por fim, deve-se declarar o método adotado para identificação da escala das variáveis latentes (fixação da variância em 1 ou seleção de indicador marcador unitário) (Jackson et al., 2009) e reportar tanto os coeficientes padronizados quanto os não padronizados (Nye, 2022).

2.5 Ajuste do Modelo

Nessa seção, a primeira diretriz metodológica consiste em evitar o uso acrítico de pontos de corte tradicionais (*cutoffs*), sejam os limites conservadores (SRMR $\leq 0,06$, RMSEA $\leq 0,06$ e CFI $\geq 0,95$) ou os lenientes (RMSEA $\leq 0,08$, CFI $\geq 0,90$ e SRMR $\leq 0,08$). A superação de um conjunto fixo de limiares não deve constituir o critério primário para validação de um modelo estrutural (Xia & Yang, 2019). Esses pontos de corte derivam de simulações com configurações altamente restritas (modelos com três fatores, 15 itens e cargas uniformes entre 0,70 e 0,80) (West et al., 2023), mostrando-se sensíveis ao número de indicadores, aos graus de liberdade, à magnitude dos parâmetros, ao tipo de desajuste e ao padrão de dados omissos (Groskurth et al., 2024; Niemand & Mai, 2018; West et al., 2023).

Todos os índices globais possuem potencialidades e restrições analíticas (West et al., 2023). Recomenda-se examinar os principais índices (SRMR, RMSEA e CFI) de forma combinada, evitando julgamentos isolados (Mai et al., 2021; Niemand & Mai, 2018). Diante de indícios de desajuste, desvios moderados em relação aos cortes convencionais justificam investigação diagnóstica aprofundada, distinguindo distorções decorrentes de idiossincrasias amostrais ou limitações métricas dos índices da má especificação do modelo (Nye, 2022). O ajuste exato reflete a ausência de discrepância entre a matriz implicada pelo modelo e a matriz populacional; já o ajuste aproximado admite resíduos amostrais e avalia sua magnitude prática. Assim, a rejeição do ajuste exato não implica

necessariamente inadequação teórica do modelo, dependendo o diagnóstico da relevância teórica das relações omitidas (McNeish & Wolf, 2024).

Nesse contexto, convém interpretar os índices convencionais de ajuste global como estimativas de tamanho de efeito, conforme concebidos originariamente (McNeish & Wolf, 2023), e não como testes de hipótese dicotômicos de aceitação ou rejeição (McNeish, 2023b). O SRMR exibe estabilidade superior em uma ampla diversidade de condições amostrais (Mai et al., 2021), preservando seu comportamento sob desvios de normalidade e variações no número de categorias dos itens (Maydeu-Olivares et al., 2017; Shi et al., 2020; Shi & Maydeu-Olivares, 2020). Adicionalmente, o binômio SRMR/CFI mostra-se suficiente para calibrar os erros Tipo I e Tipo II, ao passo que a proliferação desregulada de índices tende a inflar a taxa de erro Tipo I (Mai et al., 2021; Niemand & Mai, 2018).

Os pontos de corte dinâmicos (*Dynamic Fit Index*, DFI) adaptam a avaliação dos índices às peculiaridades estruturais do modelo e da amostra sob investigação. A fundamentação do DFI retoma o desenho experimental de simulação: geram-se réplicas sob condições teóricas ideais e sob cenários de má especificação, ajusta-se o modelo aos conjuntos simulados e derivam-se as distribuições empíricas dos índices de ajuste. Os limiares de corte são calibrados especificamente para discriminar entre estruturas adequadas e mal especificadas. A natureza dinâmica do método reside no recálculo contextualizado dessas referências para cada configuração amostral, superando a transposição descontextualizada de valores universais (McNeish & Wolf, 2023; McNeish & Wolf, 2024).

O arcabouço do DFI abrange metodologias para modelos unifatoriais e multifatoriais sob dados não normais e itens ordinais tratados como contínuos (McNeish, 2023a; McNeish & Wolf, 2023). Nas variantes fundamentadas em caminhos omitidos (*omitted paths DFI*, OP-DFI), a má especificação é simulada adicionando-se relações à estrutura geradora e omitindo-as no modelo de teste. Essa abordagem requer algoritmos customizados para cada configuração e pode impor penalidades desiguais ao mesmo nível nominal de desajuste em matrizes de ordens distintas (McNeish & Wolf, 2024).

Para contornar essas limitações, McNeish & Wolf (2024) formularam o DFI por discrepância direta (*direct discrepancy DFI*, 3DFI). O procedimento toma a matriz de correlações implicada pelo modelo estimado e introduz perturbações estocásticas para gerar distribuições populacionais não reproduzíveis pelo modelo. A amplitude dessas perturbações é regulada pela discrepância média absoluta (*Mean Absolute Discrepancy*, MAD), calculada como a média dos módulos dos resíduos fora da diagonal principal. A cada replicação, sorteia-se uma nova matriz de resíduos orientada pelo MAD selecionado, ajustando-se o modelo alvo aos dados gerados sob condições compatíveis e sob má especificação para derivar empiricamente as curvas de corte.

Essa metodologia proporciona três vantagens centrais para o diagnóstico do ajuste: dispensa a presunção *a priori* sobre quais parâmetros estariam omitidos; aplica-se uniformemente a variadas estruturas de covariância (como AFC clássica, modelos bifator e mediação); e expressa o desajuste em uma escala proporcional à dimensão da matriz, viabilizando comparações diretas de magnitude entre modelos. Dessa forma, a calibração acomoda um espectro amplo de perturbações plausíveis. A formulação proposta destina-se a modelos de grupo único com estrutura de médias nula ou saturada (McNeish & Wolf, 2024).

O 3DFI permite ainda incorporar não normalidade empírica, frequências observadas de categorias e dados omissos, combinando-se com estimadores robustos como o MLR

(McNeish & Wolf, 2024). A parametrização do MAD atua analogamente ao tamanho de efeito na AP. Os valores de 0,038, 0,050 e 0,060 propostos por McNeish & Wolf (2024) constituem propostas para ajuste próximo, razoável e medíocre, respectivamente, calibrando as distribuições empíricas de CFI e RMSEA para a arquitetura particular do modelo sob exame.

Outras alternativas contemporâneas abrangem os pontos de corte flexíveis (*Flexible Cutoffs*, FCO) e os testes de equivalência (*Equivalence Testing*, EQT). O FCO simula distribuições amostrais estritamente consistentes com o modelo estimado e adota as caudas empíricas como fronteiras para assinalar desajuste (Mai et al., 2021; Niemand & Mai, 2018). Embora dispense a modelagem explícita de erros estruturais, o FCO restringe o foco ao afastamento do ajuste perfeito, sem quantificar a tolerância ao erro residual (McNeish & Wolf, 2024). Por sua vez, o EQT inverte a presunção do teste de hipótese: assume *a priori* que o desajuste excede uma margem de tolerância e conclui pelo ajuste aceitável quando essa premissa é formalmente rejeitada. Sua formulação analítica dispensa simulações computacionais (Marcoulides & Yuan, 2017; Yuan et al., 2016). No entanto, os patamares de equivalência ainda ancoram-se em classificações convencionais de ajuste, variando sua sensibilidade conforme os graus de liberdade do modelo (McNeish & Wolf, 2024).

Em complemento às práticas a serem evitadas, desaconselha-se o emprego do alfa de Cronbach como indicador isolado de consistência interna (Bell et al., 2023; Cho, 2022; T. J. Dunn et al., 2014; Flora, 2020; Goodboy & Martin, 2020; Green & Yang, 2015; Hayes & Coutts, 2020; Kalkbrenner, 2023; McNeish, 2018; Trizano-Hermosilla & Alvarado, 2016), bem como a avaliação do teste de χ^2 desprovido de correções para não normalidade (Crede & Harms, 2019).

A inspeção deve incorporar índices de ajuste local: magnitudes de cargas fatoriais, cargas cruzadas, resíduos de correlação padronizados ($> 0,10$), índices de modificação (MI $> 3,84$) e mudanças esperadas nos parâmetros (*Expected Parameter Change*, EPC) (Nye, 2022). Esses indicadores devem ser interpretados em articulação com o conteúdo dos itens. Como métricas confiáveis de consistência da medida no contexto da AFC, preconiza-se o cálculo do ômega de McDonald e do limite inferior de confiabilidade (*Greatest Lower Bound*, GLB) (Bell et al., 2023; Cho, 2022; T. J. Dunn et al., 2014; Flora, 2020; Goodboy & Martin, 2020; Green & Yang, 2015; Hayes & Coutts, 2020; Kalkbrenner, 2023; McNeish, 2018; Trizano-Hermosilla & Alvarado, 2016).

No relato dos resultados, cabe reportar o χ^2 escalonado e os respectivos graus de liberdade (df) (Crede & Harms, 2019), acompanhados minimamente de RMSEA, CFI e SRMR computados com as devidas correções robustas (Nye, 2022). Recomenda-se reportar o coeficiente ômega de McDonald, idealmente complementado pelo GLB (Trizano-Hermosilla & Alvarado, 2016). Devem-se explicitar todas as decisões adotadas no refinamento do modelo (Flake & Fried, 2020), fundamentando-as na análise combinada das medidas de ajuste local e reportando expressamente as magnitudes de MI e das correlações residuais (Crede & Harms, 2019).

2.6 Comparações e Modificações dos Modelos

Conforme enfatizado na fase inicial, recomenda-se planejar a investigação de estruturas fatoriais alternativas desde a concepção do estudo, procedendo a sua comparação formal. Embora o pesquisador possa privilegiar determinado modelo teórico, sua escolha deve ser respaldada pelo confronto sistemático com representações conceituais concorrentes

(Jackson et al., 2009). Uma vez que nenhum modelo empírico reflete perfeitamente a realidade populacional, as estruturas fatoriais devem ser tratadas como aproximações parcimoniosas (Preacher & Yaremych, 2023).

Para que o conhecimento avance nas ciências sociais, é indispensável que teorias rivais sejam confrontadas empiricamente ao longo do tempo. Diversos atributos caracterizam um modelo de medida consistente, como interpretabilidade conceitual, simplicidade paramétrica e estabilidade preditiva. No entanto, o critério determinante reside na demonstração de desempenho superior frente a formulações alternativas em aspectos relevantes. Desse modo, o modelo retido opera como a melhor hipótese de trabalho provisória (Preacher & Yaremych, 2023).

Na seleção de modelos, o pesquisador ajusta formulações alternativas à mesma base de dados, preservando idênticos procedimentos de estimação e rotinas computacionais. Essa padronização mitiga a discricionariedade analítica e restringe o espaço para manipulação deliberada dos resultados, promovendo maior rigor epistemológico na investigação empírica (Preacher & Yaremych, 2023).

A comparação de modelos identifica a estrutura teórica mais aderente aos dados, permitindo descartar especificações deficientes e reter formulações promissoras para investigações subsequentes (Preacher & Yaremych, 2023). Assim, a seleção sistemática de modelos configura um procedimento estruturante para a consolidação de teorias psicométricas transparentes e rigorosas (Preacher & Yaremych, 2023).

Caso o modelo original apresente desajuste substancial, modificações estruturais podem ser ponderadas, desde que respaldadas em justificativas teóricas sólidas ou em evidências psicométricas prévias documentadas na literatura (Nye, 2022). A presunção de que nenhum ajuste possa ser realizado em um instrumento validado mostra-se danosa na prática: problemas de especificação na mensuração propagam-se para os parâmetros estruturais da modelagem (Brown, 2023), enquanto intervenções criteriosas na matriz de medida resguardam a estabilidade das relações inerentes ao modelo (Flake et al., 2017; Jackson et al., 2009).

No entanto, desaconselham-se alterações profundas ou arbitrárias nos modelos de medida (Brown, 2023). Na vigência de desajustes severos e generalizados, convém recorrer a uma etapa exploratória preliminar ou avaliar instrumentos de coleta alternativos. Por outro lado, caso o desajuste decorra exclusivamente da restrição de nulidade das cargas secundárias imposta pela AFC clássica, a modelagem de equações estruturais exploratória (ESEM, do inglês *Exploratory SEM*) representa uma estratégia eficiente para acomodar cargas cruzadas de modo formal (Crede & Harms, 2019).

Teoricamente, modificações empíricas sucessivas permitem alcançar o ajuste exato aos dados (modelo saturado), porém ao custo da generalização científica. Ajustes orientados estritamente por índices estatísticos alteram a natureza do estudo de confirmatória para exploratória, elevando o risco de capitalização sobre o acaso (*capitalization on chance*) (West et al., 2023). A liberação de covariâncias entre resíduos só se justifica diante de fundamentação conceitual e metodológica, como conteúdos semânticos redundantes ou polaridades linguísticas idênticas entre pares de itens (Brown, 2023; Nye, 2022). Para salvaguardar a estabilidade do modelo reespecificado diante de buscas *post hoc*, a literatura recomenda estratégias de validação cruzada (*cross-validation*), seja por meio de amostras independentes ou pela partição aleatória da base de dados (*split-sample*) em subamostras de calibração e validação (Preacher & Yaremych, 2023).

Nesse sentido, toda e qualquer alteração na especificação do modelo de medida deve ser documentada de forma transparente (Flake & Fried, 2020). Embora o ideal seja prever eventuais ajustes *a priori* (Flake & Fried, 2020), na prática psicométrica reespecificações são frequentes (Flake et al., 2017; Nye, 2022), residindo o rigor científico na transparência do relato (Flake et al., 2022; Flake & Fried, 2020). Exclusões de itens devem ser detalhadas a partir dos diagnósticos psicométricos da pesquisa, e qualquer correlação residual adicionada deve vir acompanhada de sua justificativa teórica e metodológica (Nye, 2022). Não se sustenta a alegação de restrição de espaço editorial: o pesquisador pode disponibilizar *scripts*, matrizes e saídas analíticas completas em repositórios abertos permanentes.

No entanto, quais métricas devem ser empregadas para comparar e selecionar modelos alternativos? Os índices descritivos globais (SRMR, RMSEA e CFI) foram concebidos primariamente para julgar modelos isolados e possuem limitações para a seleção direta entre alternativas (Preacher & Yaremych, 2023), ainda que existam pontos de corte para contrastar variações (Δ) desses índices tanto em modelos aninhados quanto não aninhados (P.-H. Huang, 2017), inclusive sob desvios de normalidade (Lai, 2020) ou variáveis categóricas (Lai, 2021a).

Em modelos parametricamente aninhados (nos quais uma especificação decorre da imposição de restrições paramétricas sobre outra mais geral), o método formal de comparação é o teste de diferença de qui-quadrado ($\Delta\chi^2$) ajustado (Brown, 2015). Ao empregar estimadores robustos (como MLM ou MLR), a subtração direta entre os valores ajustados de χ^2 não fornece a estatística de teste corrigida (Brown, 2015). Dessa forma, é imperativo calcular a diferença formal de χ^2 escalonado específica para o estimador adotado (Brown, 2015).

Para comparar modelos alternativos não aninhados na estrutura paramétrica, a literatura recomenda o emprego de critérios de informação, como o AIC (*Akaike Information Criterion*) e o BIC (*Bayesian Information Criterion*) (Preacher & Yaremych, 2023). Esses critérios ponderam o ajuste dos dados frente à penalização pela complexidade paramétrica a partir da função de verossimilhança da estimação por ML, herdando premissas dessa abordagem (Preacher & Yaremych, 2023). Sob dados omissos tratados via FIML, a interpretação exige cuidado redobrado, pois padrões de ausência combinados a erros de especificação podem favorecer soluções distintas daquelas identificadas com bases completas (Lai, 2021b).

Cabe destacar, no entanto, uma premissa matemática fundamental frequentemente negligenciada no uso de critérios de informação: AIC e BIC são válidos exclusivamente para comparar modelos estimados sobre o **exato mesmo conjunto de variáveis observadas** (Preacher & Yaremych, 2023). A função de verossimilhança conjunta e o valor de discrepância F_{ML} dependem diretamente do número de variáveis manifestas (p) e da ordem da matriz de covariâncias analisada (Preacher & Yaremych, 2023). Portanto, quando um modelo é modificado pela exclusão de indicadores (por exemplo, eliminando itens com baixas cargas ou resíduos elevados), o espaço amostral é alterado e as funções de verossimilhança tornam-se matematicamente incomensuráveis (Preacher & Yaremych, 2023). Não se podem empregar AIC ou BIC para afirmar que uma escala reduzida apresenta ajuste superior ao da estrutura original completa. Nesses cenários de refinamento de itens, a decisão deve pautar-se no julgamento teórico sobre a representatividade do domínio do construto e na inspeção dos resíduos locais, reservando-se a dados independentes o teste da estabilidade da estrutura enxuta (Preacher &

Yaremych, 2023).

Quando os modelos alternativos compartilham as mesmas variáveis observadas (como no confronto entre uma estrutura unifatorial, fatores correlacionados ou modelo bifator), o AIC e o BIC operam com pleno respaldo formal (Preacher & Yaremych, 2023). Em amostras moderadas, a penalização do AIC favorece a eficiência preditiva assintótica, enquanto o BIC impõe penalidade mais rigorosa pelo número de parâmetros livres, priorizando a consistência e modelos parcimoniosos (P.-H. Huang, 2017; Preacher & Yaremych, 2023). Como consistência e eficiência assintótica não podem ser otimizadas simultaneamente, é prudente reportar e contrastar ambos os critérios (Preacher & Yaremych, 2023), complementando a análise com procedimentos de validação cruzada preditiva (*k-fold*) (Preacher & Yaremych, 2023).

Nas situações em que a validação permite utilizar partição amostral, a estabilidade das estimativas pode ser examinada por meio da AFC multigrupo, tratando calibração e validação como grupos distintos e preservando o mesmo conjunto de indicadores. A análise avalia a equivalência paramétrica do modelo retido entre as subamostras. A proximidade dos índices globais entre grupos é informativa, mas não assegura a invariância de cargas, interceptos ou variâncias de erro (Fischer & Karl, 2019; Putnick & Bornstein, 2016).

A literatura estabelece quatro níveis hierárquicos de invariância. A invariância configural atesta a identidade do padrão de fatores e cargas livres entre os grupos; a invariância métrica impõe a igualdade das cargas fatoriais não padronizadas; a invariância escalar adiciona a equivalência dos interceptos; e a invariância estrita exige a igualdade das variâncias residuais (Putnick & Bornstein, 2016). O modelo configural é estimado simultaneamente nos grupos, assegurando sua identificação empírica. As restrições sucessivas geram modelos aninhados, permitindo avaliar a sustentabilidade empírica de cada nível por meio de testes de diferença de χ^2 e variações nos índices de ajuste (Fischer & Karl, 2019).

Em uma partição aleatória proveniente da mesma base de dados, calibração e validação configuram frações da mesma população de origem. Quando a finalidade consiste em verificar se as relações estruturais se mantêm entre as parcelas, a confirmação da invariância configural e métrica responde à questão central de estabilidade (Fischer & Karl, 2019). A progressão para o nível escalar torna-se mandatória apenas quando se pretende comparar médias latentes entre grupos, enquanto o nível estrito investiga a identidade dos resíduos (Putnick & Bornstein, 2016). Tais extensões detalham as propriedades da medida, mas não constituem pré-requisitos para assegurar a estabilidade das cargas fatoriais (Fischer & Karl, 2019; Putnick & Bornstein, 2016).

O teste de diferença de qui-quadrado ($\Delta\chi^2$), devidamente corrigido para o estimador robusto adotado, pode ser acompanhado pelas variações de CFI e RMSEA (ΔCFI e $\Delta RMSEA$) (Chen, 2007; Cheung & Rensvold, 2002). Considerando o modelo restrito em comparação ao precedente, variações de $\Delta CFI \leq -0,010$ e de $\Delta RMSEA \geq 0,015$ são balizas convencionais para assinalar perda de equivalência. Sua interpretação deve ser conjunta e contextualizada ao tamanho e equilíbrio das subamostras (Chen, 2007).

Uma referência empírica complementar pode ser obtida por procedimentos de permutação de grupo (amostra). O reordenamento aleatório dos participantes e a reestimação dos modelos geram uma distribuição de diferenças sob a hipótese de invariância. O valor-p empírico situa a alteração observada na cauda da distribuição nula, atenuando a

dependência de pontos de corte fixos (Jorgensen et al., 2018). A permutação deve preservar as restrições originais do desenho amostral (como a estratificação). A não rejeição das restrições fornece evidências compatíveis com a equivalência métrica do instrumento entre as subamostras (Jorgensen et al., 2018; Putnick & Bornstein, 2016).

Considerando que a seleção de modelos em SEM constitui um campo em constante desenvolvimento (Preacher & Yaremych, 2023), a premissa central é que a tomada de decisão deve estar ancorada no julgamento teórico do pesquisador (Preacher & Yaremych, 2023). Desprovidas de fundamentação teórica-conceitual, métricas puramente estatísticas tornam-se estéreis. A teoria deve delimitar *a priori* o conjunto restrito de modelos admissíveis, orientar a busca pela parcimônia conceitual durante a seleção e balizar a interpretação final das relações empíricas estimadas (Preacher & Yaremych, 2023).

A Figura 1 antecipa os procedimentos desenvolvidos nas próximas seções e suas interconexões. O mapa permite localizar cada etapa analítica e acompanhar como as especificações teóricas, a preparação dos dados e os diagnósticos estatísticos sustentam as decisões metodológicas sobre os modelos.

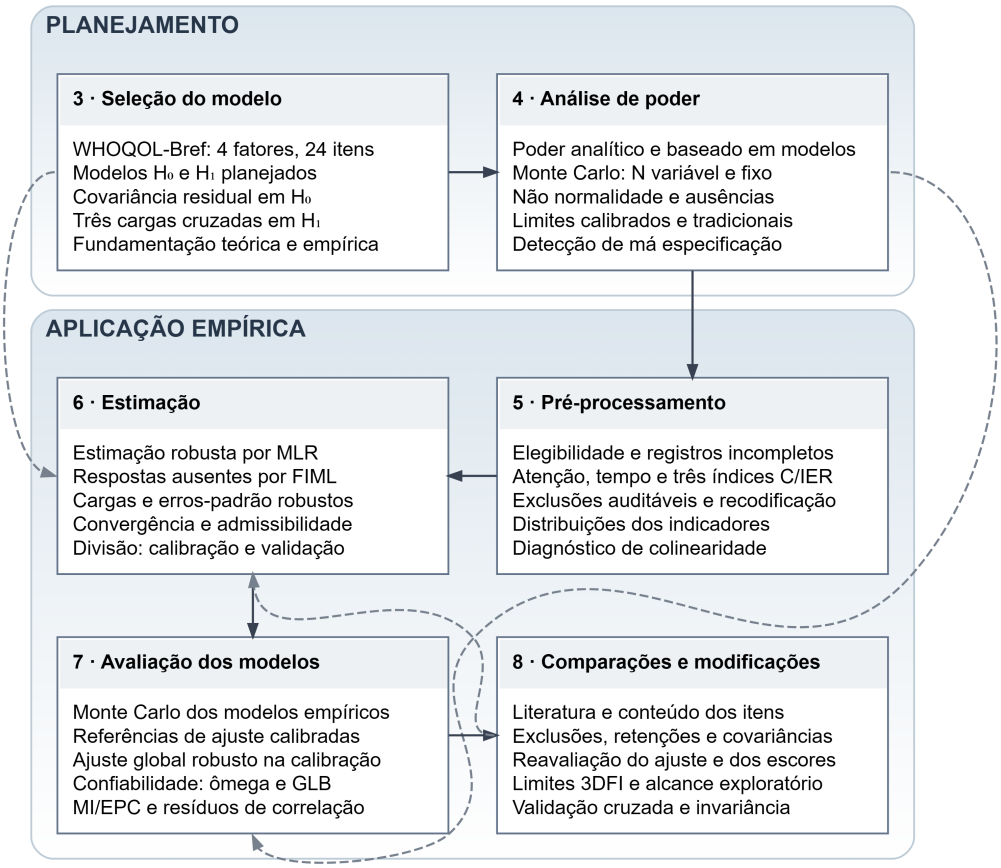


Figura 1: Mapa das decisões, procedimentos e evidências desenvolvidos na aplicação de AFC. As setas contínuas indicam o percurso de leitura; as tracejadas conectam as especificações planejadas à estimação, a calibração por simulação à avaliação empírica e as modificações à reestimação.

3 Seleção do modelo de medida

Utiliza-se como aplicação empírica a escala *World Health Organization Quality of Life - Abbreviated Version* (WHOQOL-Bref), um instrumento de autorrelato amplamente empregado na mensuração da qualidade de vida em diversas populações e contextos transculturais (Lin & Yao, 2022; Mosqueira Taipe et al., 2026; Skevington & Epton, 2018). O instrumento compreende 26 itens estruturados em cinco categorias de resposta. Os dois primeiros itens mensuram a percepção global da qualidade de vida e a satisfação geral com a saúde; os 24 itens restantes compõem quatro domínios específicos: físico (FIS), psicológico (PSI), relações sociais (SOC) e ambiente (AMB).

A escolha da escala WHOQOL-Bref operacionaliza a diretriz metodológica delineada na Seção 2.1: partir de um instrumento com fundamentação teórica consolidada e evidências psicométricas acumuladas, cujos itens viabilizem o tratamento contínuo das variáveis observadas. A presença de cinco categorias de resposta oferece sustentação empírica

inicial a essa estratégia, cujas implicações serão examinadas detalhadamente na descrição dos indicadores (Seção 5.3) e na estimação robusta (Seção 6). O propósito da aplicação consiste em explicitar as decisões de uma AFC e seus fundamentos, preservando o conteúdo dos itens e o formato métrico original do instrumento.

As sínteses da literatura corroboram a estrutura de quatro domínios, indicando simultaneamente que sua aplicação empírica requer atenção às inter-relações entre itens específicos. Lin & Yao (2022) integraram achados de 16 estudos de análise fatorial exploratória e sustentaram a solução de quatro fatores por sua interpretabilidade e parcimônia. Nessa estrutura fatorial, o item Q8 (segurança na vida diária) exibiu carga mais pronunciada no domínio psicológico do que no ambiental, ao qual se vincula teoricamente. Esse resultado evidencia que a sustentação da organização tetrafatorial pode coexistir com complexidades empíricas locais que se afastam de uma estrutura fatorial simples.

A revisão sistemática conduzida por Mosqueira Taipe et al. (2026) amplia essa constatação: a partir de 439 registros mapeados, foram incluídos 36 estudos empíricos, dos quais 64% identificaram a estrutura original de quatro fatores. A revisão respalda a arquitetura conceitual da escala, mas documenta reespecificações associadas a especificidades socioculturais e clínicas, além de expressiva heterogeneidade nos protocolos analíticos e no reporte dos índices de ajuste. Dessa forma, esse conjunto de evidências justifica adotar os quatro domínios teóricos como ponto de partida e prever formalmente as relações residuais e cruzadas a serem avaliadas.

Nesse sentido, a partir dos achados de Lin & Yao (2022), especifica-se o modelo de medida primário (H_0), que incorpora uma covariância residual entre o item Q3 (dor física e desconforto) e o item Q4 (dependência de tratamento médico) (Figura 2). Ambos integram o domínio físico e possuem enunciados direcionados negativamente à qualidade de vida. A sobreposição semântica e a polaridade compartilhada tornam plausível uma associação residual não explicada unicamente pelo fator latente físico. Essa especificação encontra precedentes nos estudos compilados por Mosqueira Taipe et al. (2026). No âmbito do viés de método comum (*Common Method Bias*, CMB), interpreta-se essa covariância residual como uma hipótese de dependência local decorrente de variância de método compartilhada ou especificidade de conteúdo.

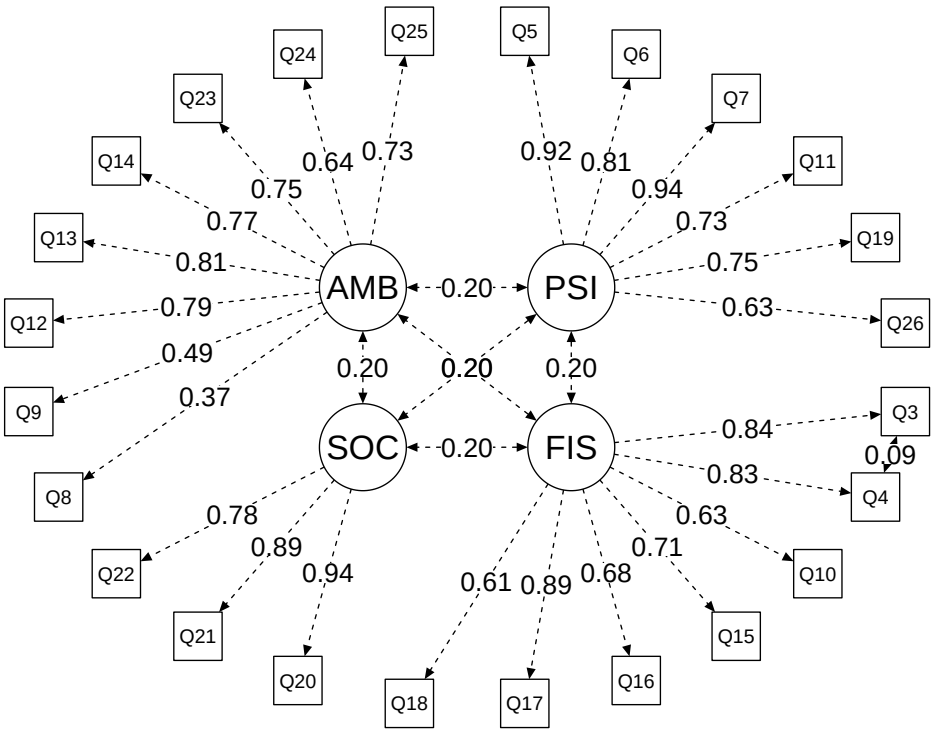


Figura 2: Modelo de quatro fatores correlacionados com covariância residual entre Q3 e Q4. *Nota.* PSI = psicológico; FIS = físico; SOC = relações sociais; AMB = ambiente. A figura representa a especificação (cargas fixas) do modelo populacional. Ao estimar com parâmetros livres, as cargas e a covariância residual são ajustadas à amostra de análise, configurando o modelo de medida primário (H_0).

O segundo modelo mantém a covariância residual entre Q3 e Q4 e adiciona três cargas cruzadas: Q8 e Q9 associados ao domínio psicológico e Q15 associado ao domínio ambiental (Figura 3). A especificação dessas relações fundamenta-se na solução fatorial exploratória documentada por Lin & Yao (2022). A segurança pessoal percebida em Q8 e as condições do espaço físico em Q9 compartilham conteúdo com o bem-estar psicológico, enquanto a mobilidade funcional aferida por Q15 envolve tanto capacidade física quanto acessibilidade ambiental. Essas interpretações fornecem justificativas teóricas para testar tais parâmetros, constituindo hipóteses a serem confrontadas empiricamente na amostra do presente tutorial.

Mantêm-se as cargas principais de Q8 e Q9 no domínio ambiental e a de Q15 no domínio físico, acrescentando-se as relações propostas em vez de reatribuir os itens de domínio. Tampouco se incorporam todas as cargas secundárias da solução exploratória de Lin & Yao (2022). O modelo alternativo (mal especificado) configura uma especificação delimitada para o presente estudo, apoiada na literatura empírica e na análise semântica dos itens, e não uma AFC previamente validada com essa exata parametrização.

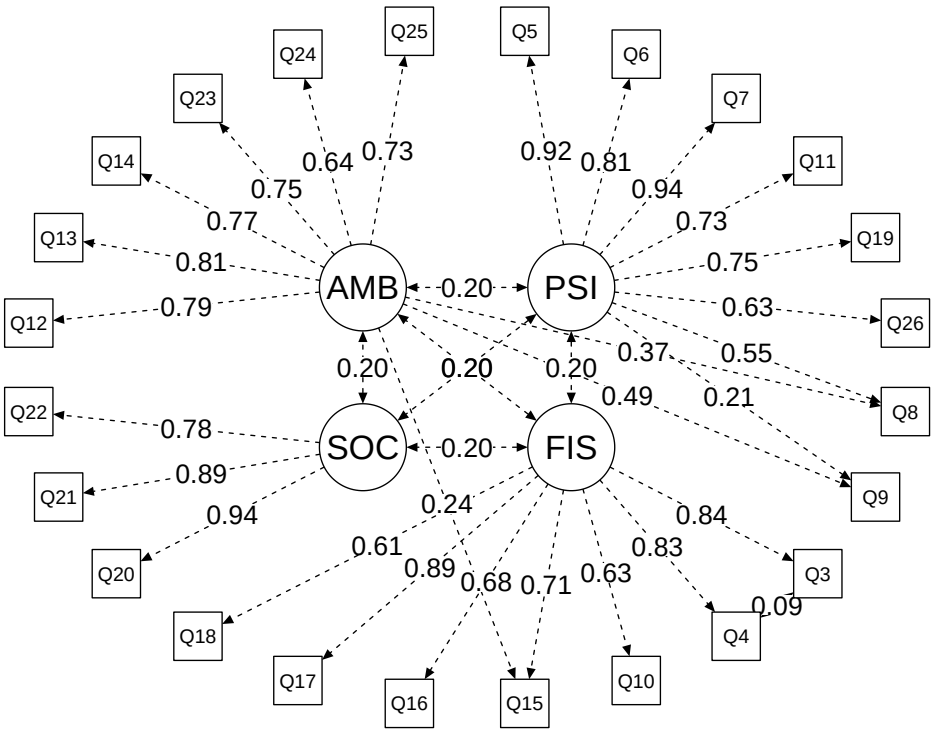


Figura 3: Modelo de quatro fatores correlacionados com covariância residual entre Q3 e Q4 e três cargas cruzadas. *Nota.* PSI = psicológico; FIS = físico; SOC = relações sociais; AMB = ambiente. Mantém-se a estrutura do modelo primário e acrescentam-se as cargas do fator psicológico em Q8 e Q9 e do fator ambiental em Q15. A figura representa a especificação (cargas fixas) do modelo populacional alternativo. Ao estimar com parâmetros livres, as cargas e a covariância residual são ajustadas à amostra, configurando o modelo alternativo (H_1).

4 Análise de poder

Em pesquisas empíricas, a escala psicométrica atua frequentemente como modelo de mensuração de construtos latentes a serem integrados em um modelo estrutural posterior. O planejamento amostral deve fundamentar-se no instrumento específico e na estrutura fatorial a ser estimada, superando regras heurísticas genéricas sobre o número de itens ou parâmetros livres (Kline, 2023). Nessa seção, define-se o poder estatístico como a probabilidade de detectar uma omissão estrutural formalmente delimitada antes da coleta de dados.

4.1 Pressupostos a priori

O planejamento da AP ancora-se nas duas especificações do modelo de medida fundamentadas na Seção 3. Nessa análise, o segundo modelo representa a condição populacional

alternativa na qual a adoção da estrutura parcimoniosa implica a omissão deliberada de três cargas cruzadas.

As cargas fatoriais principais, as correlações inter-fatores e as variâncias residuais foram parametrizadas com base nas evidências meta-analíticas e revisões sistemáticas prévias (Lin & Yao, 2022; Mosqueira Taipe et al., 2026; Skevington & Epton, 2018). Dessa forma, a AP é conduzida estritamente *a priori*: as características populacionais e a magnitude da má especificação são estabelecidas antes do acesso aos dados empíricos examinados nas seções posteriores.

Além da configuração fatorial, as simulações incorporam duas condições metodológicas frequentes na aplicação da AFC: desvios de normalidade e dados omissos. A distribuição dos indicadores contempla assimetrias entre -1 e $+1$ e curtose entre 1,7 e 2,0, refletindo desvios moderados de normalidade compatíveis com dados empíricos da escala WHOQOL-Bref (Yoshitake et al., 2015).

Admite-se, igualmente, que 10% das respostas estejam ausentes sob o mecanismo completamente ao acaso (*Missing Completely at Random*, MCAR). A adoção desse padrão modela a perda informacional sem demandar hipóteses condicionais complexas sobre variáveis auxiliares. Em todas as condições experimentais, os parâmetros são estimados por MLR, associada ao tratamento de dados omissos por FIML. A Tabela 1 sintetiza as especificações adotadas.

Tabela 1: Pressupostos da análise de poder a priori.

Componente	Especificação	Base
Modelo de referência	Quatro fatores correlacionados e covariância residual Q3–Q4	Lin e Yao (2022)
Misspecificação	Omissão de três cargas cruzadas	Mosqueira-Taipe et al. (2026)
Não normalidade	Assimetria de -1 a +1; curtose entre 1,7 e 2,0	Yoshitake et al. (2015)
Dados ausentes	10% MCAR	Mosqueira-Taipe et al. (2026)
Estimação	MLR + FIML	Cenário realístico
Erro tipo I	0.05	Definido a priori
Poder-alvo	0.80	Definido a priori

4.2 Abordagem analítica de poder

A primeira estimativa do tamanho amostral mínimo decorre da derivação analítica baseada em modelos. Nessa abordagem, a população geradora incorpora as três cargas cruzadas teóricas, ao passo que o modelo estimado restringe a análise às cargas principais e à covariância residual entre Q3 e Q4. A magnitude do efeito decorre da discrepância populacional entre a matriz implicada e as restrições estruturais impostas pelo modelo de análise (Moshagen & Bader, 2024).

Sem a especificação de um modelo alternativo aninhado estimado livremente, o teste avalia o ajuste global do modelo de análise frente ao modelo saturado. Dessa forma, a hipótese estatística quantifica a probabilidade de rejeitar formalmente a estrutura parcimoniosa diante da presença real das cargas cruzadas na população.

Adotando-se nível de significância de 5% e poder estatístico-alvo de 80%, o procedimento analítico indicou a necessidade de um tamanho amostral mínimo de 106 participantes. Esse valor opera como baliza inicial para os estudos de simulação, e não como determinação amostral definitiva.

A Tabela 2 apresenta as métricas resultantes da derivação analítica e o tamanho de amostra necessário para alcançar o poder estabelecido.

Tabela 2: Resultados da abordagem analítica de poder baseada em modelos.

Medida	Símbolo	Valor
Discrepância populacional	F_0	0.579
Raiz do erro quadrático médio de aproximação	RMSEA	0.049
Índice de centralidade de McDonald	Mc	0.749
Índice de qualidade do ajuste	GFI	0.954
Índice de qualidade do ajuste ajustado	AGFI	0.944
Raiz do resíduo quadrático médio padronizado	SRMR	0.067
Índice de ajuste comparativo	CFI	0.960
Graus de liberdade	df	245
Tamanho amostral necessário	N	106
Qui-quadrado crítico	χ^2_{crit}	282.511
Parâmetro de não centralidade	λ	60.793
Nível de significância (alfa)	α	0.050
Erro tipo II (beta)	β	0.1968
Poder estatístico	$1 - \beta$	0.8032

A Figura 4 ilustra as distribuições central e não central do qui-quadrado. A distribuição sob a hipótese nula representa o comportamento esperado da estatística sob correta especificação estrutural, ao passo que a distribuição não central projeta a dispersão diante da discrepância populacional estipulada. A área de rejeição e a sobreposição das curvas expressam graficamente a relação entre o erro Tipo I e o poder estatístico do teste.

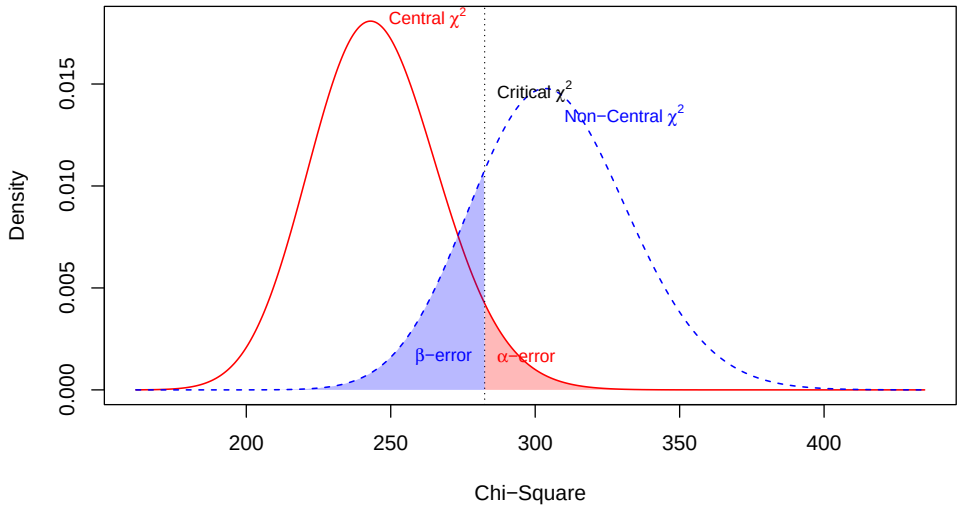


Figura 4: Distribuições central e não central do qui-quadrado da abordagem analítica de poder baseada em modelos.

4.3 SMC com tamanho amostral variável

Embora a derivação assintótica forneça uma estimativa inicial objetiva, não reproduz integralmente as condições empíricas observadas na prática da pesquisa. Dessa forma, a fase subsequente implementa uma SMC combinando desvios moderados de normalidade e 10% de dados omissos MCAR, preservando-se a estimação por MLR com tratamento via FIML.

Os tamanhos amostrais variam de 101 a 400 participantes. Para cada tamanho de amostra, geram-se dez replicações sob duas condições populacionais: na condição de referência, os dados reproduzem o modelo de quatro fatores sem cargas cruzadas; na condição alternativa, os dados incorporam as três cargas secundárias, mas são estimados pelo modelo parcimonioso. O confronto dessas condições permite estimar o poder empírico dos índices RMSEA, SRMR, CFI e TLI na identificação da má especificação planejada.

Apresentam-se duas configurações gráficas da relação entre tamanho amostral e poder estatístico. A primeira adota suavização logística para destacar a trajetória de evolução do poder ao longo do gradiente amostral (Figura 5). A segunda exibe diretamente as taxas empíricas de rejeição observadas nas replicações realizadas em cada patamar (Figura 6).

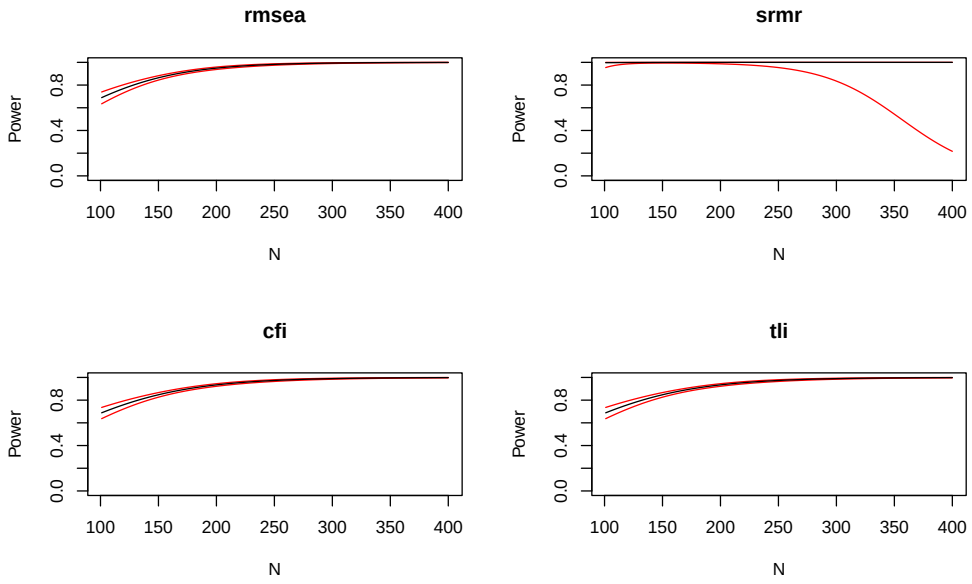


Figura 5: Curvas de poder com suavização logística sob não normalidade leve e 10% de dados ausentes MCAR.

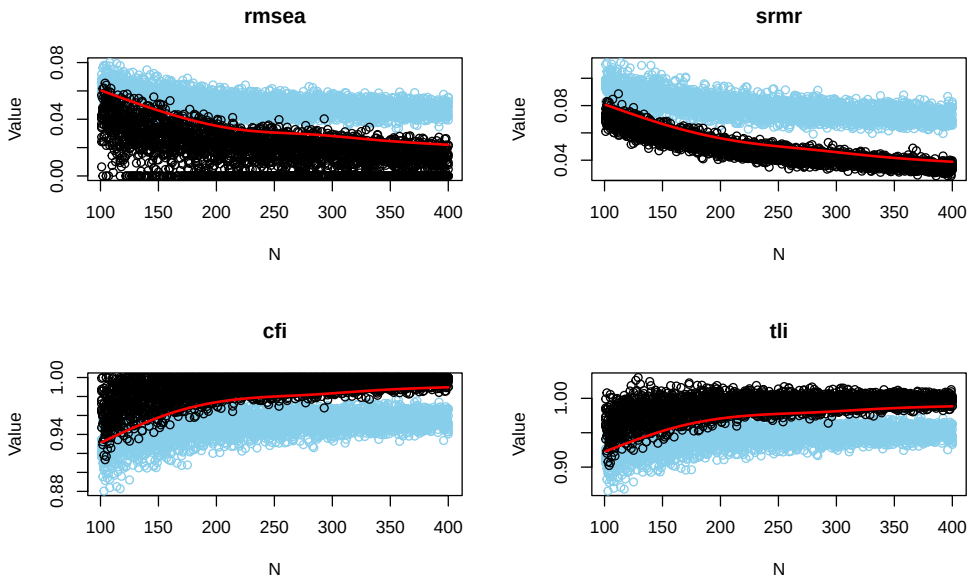


Figura 6: Poder empírico sem suavização logística sob não normalidade leve e 10% de dados ausentes MCAR.

As curvas apresentadas na Figura 5 e na Figura 6 indicam que uma amostra próxima a

150 respondentes atinge sensibilidade suficiente para que os índices globais identifiquem a omissão das cargas cruzadas. Calculam-se, então, o poder e os pontos de corte empíricos especificamente para $N = 150$. Essa verificação permite aferir se o patamar indicado visualmente pelas curvas mantém estabilidade entre os quatro índices globais ou se opera excessivamente próximo do limiar crítico.

Tabela 3: Poder e pontos de corte calibrados pelo modelo para N = 150.

	RMSEA	SRMR	CFI	TLI
Poder	0.864	0.999	0.847	0.847
Ponto de corte	0.046	0.068	0.957	0.951

Em $N = 150$, o poder estimado atingiu 0.864 para o RMSEA, 0.999 para o SRMR, 0.847 para o CFI e 0.847 para o TLI.

Essa etapa inicial avalia especificamente a sensibilidade dos índices globais de ajuste à má especificação estrutural. Cumpre ressalvar que a adequação amostral para o teste de ajuste global não assegura, isoladamente, a precisão assintótica na recuperação pontual de parâmetros ou na cobertura de intervalos de confiança (Feng & Hancock, 2023; Pornprasertmanit et al., 2022), justificando cautela na fixação do tamanho amostral definitivo.

4.4 SMC com tamanho amostral fixo

Considerando os resultados das curvas de poder e a necessidade de resguardar a precisão das estimativas sob desvios de normalidade e dados omissos, adota-se $N = 200$ como tamanho amostral prudencial de planejamento no presente tutorial.

A análise com tamanho amostral fixo emprega 1.000 replicações para cada condição populacional. Na primeira, simulam-se dados consistentes com o modelo de referência; na segunda, os dados incorporam as três cargas cruzadas, mantendo-se a estimação pelo modelo parcimonioso. Diferentemente da fase anterior, na qual dez replicações foram distribuídas em múltiplos tamanhos amostrais, essa etapa concentra todas as 1.000 replicações em $N = 200$. Essa concentração assegura estimativas mais estáveis do poder estatístico e das distribuições empíricas dos índices de ajuste para esse patamar específico.

Com o tamanho amostral constante, dispensa-se a interpolação logística. A Figura 7 retrata diretamente o comportamento empírico dos índices nas duas condições populacionais. A separação entre as curvas ilustra a capacidade do teste em evidenciar a má especificação estrutural.

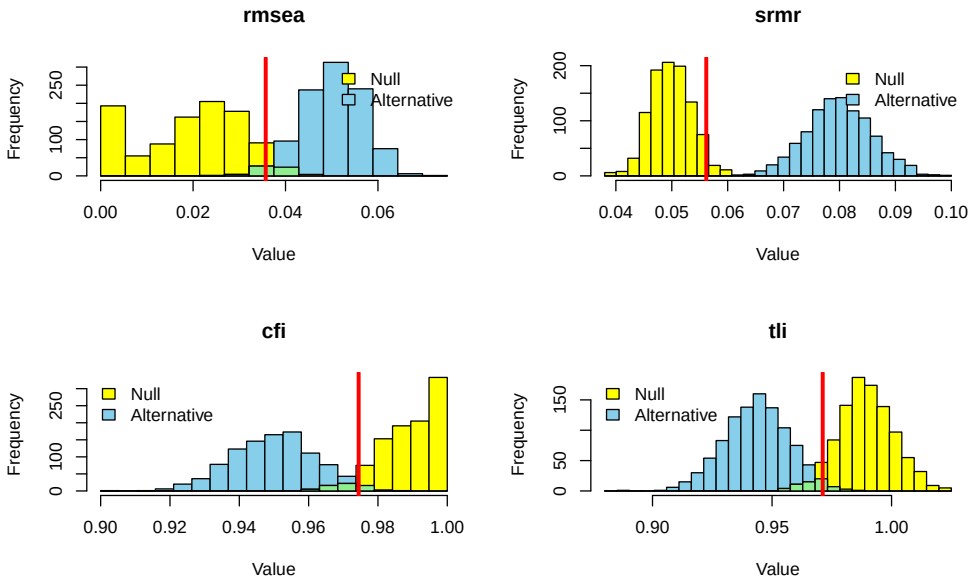


Figura 7: Distribuições empíricas dos índices de ajuste em 1.000 replicações com N = 200.

Calculam-se, em seguida, o poder de detecção da má especificação e os pontos de corte correspondentes a RMSEA, SRMR, CFI e TLI. Esses limiares são calibrados a partir da distribuição empírica sob correta especificação, integrando simultaneamente tamanho amostral, desvios de normalidade, taxa de omissão e método de estimação. A Tabela 4 reúne os resultados para $N = 200$.

Tabela 4: Poder e pontos de corte calibrados pelo modelo em 1.000 replicações com N = 200.

	RMSEA	SRMR	CFI	TLI
Poder	0.984	1.000	0.982	0.982
Ponto de corte	0.036	0.056	0.974	0.971

4.5 Comparação com pontos de corte tradicionais

Por fim, compara-se o desempenho dos pontos de corte calibrados empiricamente com os limiares heurísticos tradicionais de ajuste global. Para RMSEA e SRMR, a rejeição ocorre quando as estimativas ultrapassam os limites teóricos estabelecidos; para CFI e TLI, verifica-se a rejeição quando os valores situam-se abaixo dos respectivos cortes.

A simulação com tamanho amostral fixo é reavaliada mantendo-se inalterados o tamanho amostral, as matrizes populacionais e as propriedades empíricas dos dados, modificando-se estritamente o critério de decisão. A Figura 8 exibe a classificação decorrente dos pontos de corte convencionais, e a Tabela 5 resume o poder resultante.

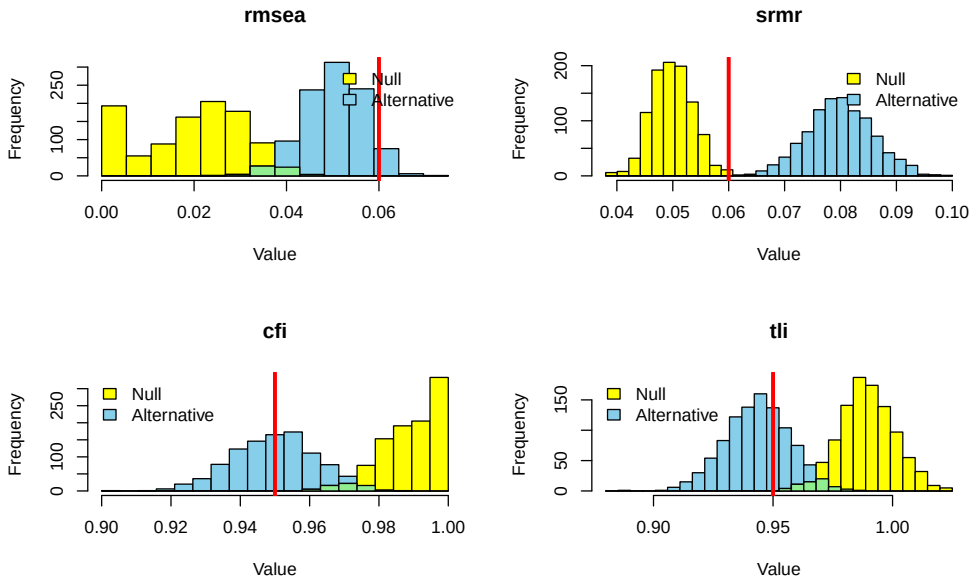


Figura 8: Distribuições dos índices de ajuste avaliadas por pontos de corte tradicionais em 1.000 replicações com N = 200.

Tabela 5: Poder obtido com pontos de corte tradicionais em 1.000 replicações com N = 200.

	CFI	TLI	RMSEA	SRMR
Poder	0.496	0.678	0.062	1

Sob os pontos de corte tradicionais, o poder observado foi de 0.062 para o RMSEA, 1.000 para o SRMR, 0.496 para o CFI e 0.678 para o TLI.

Essa comparação evidencia que as regras de bolso convencionais geram diagnósticos substancialmente divergentes dos limiares calibrados por simulação, distorcendo a sensibilidade do teste à má especificação do modelo (Groskurth et al., 2024; McNeish & Wolf, 2023).

Dessa forma, as três etapas da AP cumprem funções complementares: a abordagem analítica fornece uma estimativa inicial fundamentada na discrepância populacional; a simulação com tamanho variável descreve a evolução da curva de poder entre 101 e 400 participantes; e a simulação intensiva com 1.000 replicações em $N = 200$ confere estabilidade assintótica ao planejamento amostral adotado.

Os pontos de corte derivados em $N = 200$ vinculam-se estritamente às premissas desse planejamento: parâmetros populacionais informados pela literatura empírica, desvios moderados de normalidade e 10% de dados omissos MCAR. Sua transposição para amostras com outras dimensões, distribuições ou configurações exige calibração específica. O resultado central dessa etapa reside na determinação do poder estatístico esperado para identificar a omissão estrutural planejada sob o cenário simulado.

5 Pré-processamento

O pré-processamento em questionários de autorrelato tem como finalidade proteger as análises subsequentes contra vieses decorrentes de C/IER, registros incompletos ou respostas inconsistentes, além de inspecionar preliminarmente as variáveis para a detecção precoce de anomalias. Diferentemente das seções anteriores, em que os dados empíricos não foram demandados, as etapas subsequentes operam diretamente sobre dados brutos, adotando-se a mesma base empregada por Rogers (2024) e Rogers (2022).

No presente tutorial, no entanto, o processamento parte dos 1.546 registros brutos e explicita tanto os critérios procedurais (*ex-ante*) quanto os indicadores estatísticos (*ex-post*) de triagem C/IER. Como os trabalhos anteriores (Rogers, 2022, 2024) partiram de uma base pré-processada de 1.047 participantes, eventuais diferenças entre seus resultados e os do presente tutorial não devem ser atribuídas exclusivamente ao tratamento ordinal ou contínuo dos itens. Após a verificação da integridade do banco e a remoção de registros sem identificação ou com mais de cinco respostas ausentes na WHOQOL-Bref, a amostra elegível foi submetida a um protocolo sequencial de triagem C/IER. As respostas ausentes nos registros retidos foram preservadas para tratamento via FIML na etapa de estimação.

Para o tratamento dos indivíduos com C/IER adotou-se a abordagem de múltiplos obstáculos (*multiple-hurdle approach*) recomendada por Curran (2016) e DeSimone et al. (2015). Como os índices de desatenção mensuram manifestações comportamentais distintas (a exemplo de respostas aleatórias, padrões invariantes e preenchimento excessivamente rápido) e podem apresentar correlações baixas ou nulas entre si (Biemann et al., 2025; Goldammer et al., 2020; Meade & Craig, 2012), a literatura preconiza a aplicação combinada de múltiplos critérios de triagem (Arthur et al., 2021; Curran, 2016; Ulitzsch et al., 2022). No presente tutorial, cada critério foi tratado como uma barreira obrigatória, de modo que a falha em qualquer obstáculo determinou a exclusão do participante da amostra final (Curran, 2016; Goldammer et al., 2020).

5.1 Exclusões procedurais

No âmbito dos controles *ex-ante*, adotaram-se itens de atenção inseridos no questionário e a duração total de preenchimento da pesquisa (Curran, 2016; Ward & Meade, 2023). A falha em pelo menos dois dos três itens de atenção incluídos foi interpretada como evidência de C/IER (Chauliac et al., 2023; DeSimone et al., 2015). Adicionalmente, foram excluídos os questionários de respondentes com velocidade inferior a quatro segundos por item, limiar adotado para sinalizar preenchimentos excessivamente rápidos nessa aplicação, situando-se dentro do intervalo de referência de dois a quatro segundos por item discutido na Seção 2.3.

5.2 Critérios estatísticos

Como critérios *ex-post*, empregaram-se a medida de *longstring*, o IRV e a distância de Mahalanobis (Biemann et al., 2025; A. M. Dunn et al., 2018). O primeiro indicador permite identificar padrões mecânicos de marcação uniforme, o segundo detecta tanto variabilidade anormalmente baixa quanto dispersão atipicamente elevada, e o terceiro aponta padrões multivariados distantes do comportamento normativo da amostra, condizentes com aleatoriedade excessiva (Curran, 2016; DeSimone et al., 2015; Hong et al., 2020; Meade & Craig, 2012). Considerou-se problemática uma sequência contínua

(*longstring*) superior a 13 respostas consecutivas, o que corresponde à metade dos 26 itens da WHOQOL-Bref avaliados na triagem. Para o IRV, adotaram-se os percentis 1% e 99%, enquanto para a distância de Mahalanobis estabeleceu-se o percentil 99%. As regras de triagem foram delimitadas com base na extensão da escala e nas caudas das distribuições empíricas. Os pontos de corte foram calculados diretamente sobre a amostra submetida aos critérios estatísticos, buscando equilibrar a sensibilidade à desatenção e a preservação de respondentes sinceros (Arthur et al., 2021; Goldammer et al., 2020; Steedle, 2018).

Tabela 6: Pontos de corte dos indicadores estatísticos.

Indicador	Média	Desvio-padrão	Mediana	1º percentil	99º percentil	Regra de exclusão
Longstring	5.216	2.734	4.000	2.000	15.000	longstring > 13
IRV	1.025	0.207	1.027	0.552	1.491	IRV p1 ou IRV p99
Distância de Mahalanobis	26.000	15.336	22.060	5.844	80.639	Mahalanobis p99

5.2.1 Decisão final e amostra de análise

Após o cálculo dos indicadores de atenção, velocidade, invariância e discrepância multivariada, cada participante foi avaliado simultaneamente nos cinco obstáculos definidos no protocolo. A decisão de exclusão foi aplicada de forma estrita: a falha em qualquer barreira determinou a exclusão do respondente da amostra de análise. A Tabela 7 resume a frequência de sinalização em cada critério e permite avaliar quais manifestações de resposta descuidada foram mais frequentes na amostra elegível. As sinalizações associadas aos itens de atenção e ao IRV foram as mais recorrentes, ao passo que o tempo de resposta apresentou menor incidência. Essa distribuição reforça a utilidade da estratégia de múltiplos obstáculos, pois os critérios não identificam exatamente os mesmos participantes nem representam uma única modalidade de resposta inadequada.

Tabela 7: Frequência de falhas em cada obstáculo de limpeza.

Critério	N sinalizado	Percentual (%)
Itens de atenção	28	2.1
Tempo de resposta	8	0.6
Longstring	21	1.6
IRV	28	2.1
Distância de Mahalanobis	14	1.1

A aplicação conjunta dos cinco obstáculos resultou em redução amostral de 1.546 para 1.243 participantes. Dos 1.546 registros brutos, 223 foram eliminados na etapa preliminar por não apresentarem as condições mínimas de preenchimento da escala. Entre os 1.323 respondentes elegíveis aos obstáculos, 80 falharam em pelo menos um critério e foram excluídos da base analítica. Essa trajetória evidencia que a maior perda amostral ocorreu

antes da aplicação dos indicadores de C/IER, ao passo que a triagem comportamental produziu uma exclusão proporcionalmente menor, concentrada em padrões incompatíveis com o preenchimento atencioso. A Tabela 8 sintetiza essa evolução.

Tabela 8: Trajetória da amostra durante o pré-processamento.

Etapa	N
Banco bruto	1546
Dropouts removidos	223
Elegíveis aos obstáculos	1323
Excluídos por pelo menos um obstáculo	80
Amostra final	1243

5.3 Descrição dos itens

Com a amostra de análise delimitada, a etapa seguinte consistiu em avaliar o comportamento empírico dos itens que integram o modelo de AFC. Essa inspeção preliminar verifica o volume de respostas disponíveis, a tendência central, a dispersão e o formato das distribuições, fornecendo subsídios diretos para a escolha do estimador psicométrico. Em termos gerais, os itens apresentaram médias superiores ao ponto médio da escala e variabilidade adequada para capturar as diferenças individuais. A proporção de dados ausentes variou discretamente entre os itens, mantendo-se em níveis residuais e justificando a retenção dos participantes mediante tratamento por FIML. A Tabela 9 sintetiza essas estatísticas.

Tabela 9: Estatísticas descritivas dos itens utilizados na AFC.

Item	Respostas válidas	Média	Desvio-padrão	Assimetria	Curtose	VIF
Q3	1242	4.136	0.941	-0.910	0.076	1.624
Q4	1241	4.089	1.009	-0.983	0.183	1.473
Q5	1230	3.283	0.818	-0.472	-0.008	2.140
Q6	1238	3.895	0.974	-0.871	0.534	1.752
Q7	1241	3.586	0.858	-0.489	0.059	1.706
Q8	1240	2.933	0.979	-0.098	-0.459	1.317
Q9	1239	3.471	0.884	-0.569	0.216	1.475
Q10	1241	3.587	0.903	-0.164	-0.368	2.226
Q11	1241	3.798	0.996	-0.615	-0.087	1.611
Q12	1242	3.414	1.099	-0.241	-0.530	1.825
Q13	1243	3.896	0.828	-0.570	0.235	1.545
Q14	1240	3.331	0.956	-0.281	-0.245	1.927
Q15	1243	4.398	0.807	-1.513	2.455	1.504
Q16	1232	3.485	1.035	-0.124	-1.165	1.430
Q17	1242	3.711	0.921	-0.504	-0.538	3.689
Q18	1243	3.730	0.955	-0.488	-0.659	3.247
Q19	1240	3.691	0.942	-0.400	-0.706	2.661
Q20	1239	3.608	0.891	-0.237	-0.674	1.893
Q21	1242	3.420	1.009	-0.003	-1.111	1.316
Q22	1241	3.554	0.884	-0.131	-0.703	1.730

Tabela 9: Estatísticas descritivas dos itens utilizados na AFC.

Item	Respostas válidas	Média	Desvio-padrão	Assimetria	Curtose	VIF
Q23	1237	3.973	0.949	-0.686	-0.413	1.392
Q24	1242	3.725	0.963	-0.400	-0.771	1.389
Q25	1241	3.821	0.978	-0.489	-0.737	1.402
Q26	1243	3.891	0.930	-1.160	1.535	1.559

A Tabela 9 permite também avaliar a distribuição dos itens. Observa-se assimetria predominantemente negativa, sobretudo nos itens com maior concentração nas categorias superiores, ao passo que determinados itens exibem coeficientes de curtose mais elevados. Esses achados são compatíveis com o padrão característico de respostas em escalas Likert, nas quais a amplitude discreta de categorias e a assimetria das respostas produzem afastamentos em relação à normalidade multivariada. Por essa razão, a estimação subsequente requer procedimentos dotados de correções robustas para os erros-padrão e para as estatísticas de teste (Li, 2016a, 2016b).

A presença de dados ausentes residuais em determinados itens igualmente orienta a escolha metodológica. Como os participantes não foram descartados por omissões pontuais, a análise preserva a integralidade das informações observadas ao empregar o procedimento FIML (Lai, 2021b; Lei & Shiverdecker, 2020). Dessa forma, as respostas disponíveis de cada participante contribuem para a função de verossimilhança sem necessidade de exclusão *listwise*.

Os valores de VIF complementam a inspeção univariada ao rastrear eventuais redundâncias lineares severas entre os indicadores. Todas as estimativas situaram-se confortavelmente abaixo do ponto de corte de 5 (T. Kyriazos & Poga, 2023), assegurando a ausência de multicolinearidade crítica entre os itens da AFC e garantindo a estabilidade numérica na inversão das matrizes de covariâncias e de informação.

Como exemplo, a Figura 9 ilustra a distribuição empírica das respostas nos itens do construto Relações Sociais. O gráfico evidencia simultaneamente a concentração das marcações, a baixa frequência nas categorias inferiores e a assimetria de cada indicador. Observa-se maior adensamento nas categorias 4 e 5 da escala, padrão plenamente condizente com as médias elevadas reportadas na Tabela 9.

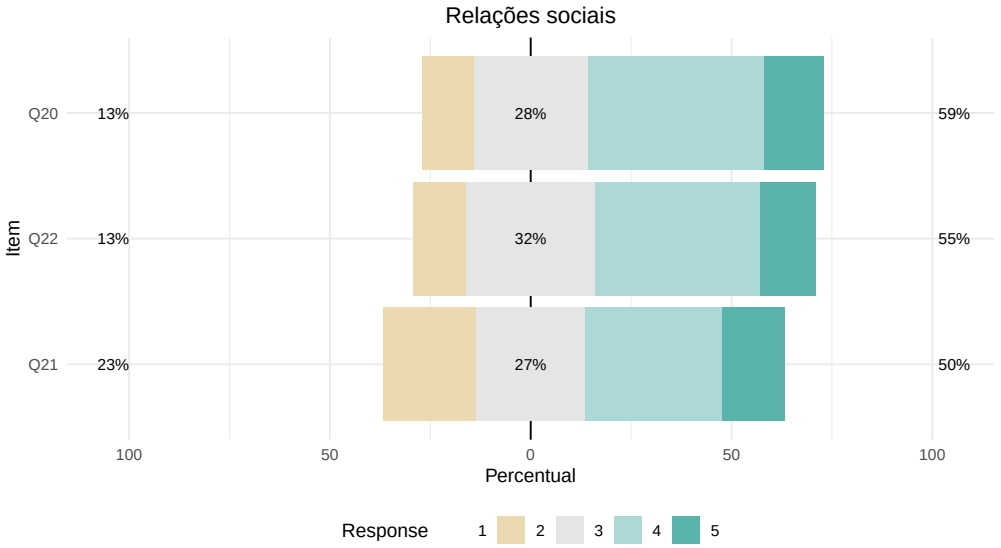


Figura 9: Distribuição das categorias de resposta Likert no construto Relações sociais.

6 Processo de estimação

Após a seleção dos modelos e o pré-processamento, a presente seção estima as duas especificações da AFC na subamostra de calibração. A estimação traduz as relações previstas nos modelos teóricos em parâmetros empíricos, sob as condições estabelecidas nas etapas anteriores: tratamento contínuo dos itens, estimador MLR, procedimento FIML e identificação pela fixação unitária da variância dos fatores (Jackson et al., 2009; Nye, 2022).

Os 1.243 participantes da amostra de análise foram divididos aleatoriamente em duas parcelas: 621 para calibração e 622 para validação. A partição foi estratificada pela presença de respostas ausentes nos 24 indicadores, distribuindo os 60 participantes com alguma omissão de forma equilibrada entre os dois grupos. A subamostra de calibração é empregada para estimar os modelos e examinar seus diagnósticos, ao passo que a subamostra de validação é reservada à confirmação empírica da reespecificação selecionada.

6.1 Fundamentação da estimação

Os itens foram tratados como variáveis contínuas, conforme a fundamentação desenvolvida na Seção 2.4. De acordo com Rhemtulla et al. (2012), o estimador MLR exibe desempenho adequado em configurações com cinco ou mais categorias ordinais, em especial quando os limiares de resposta distribuem-se de maneira aproximadamente simétrica. Robitzsch (2020) e Robitzsch (2022) demonstram que as formulações contínua e ordinal definem alvos paramétricos distintos. Com cinco categorias e uma estrutura fatorial previamente fundamentada, adota-se a formulação contínua como aproximação de mensuração no presente estudo.

O método MLR foi selecionado por combinar a ML com erros-padrão robustos e es-

tatísticas de teste corrigidas por escala. Essa combinação ajusta-se com precisão às distribuições observadas no pré-processamento e atende às recomendações metodológicas para dados sob violação de normalidade multivariada (Brown, 2015; Kline, 2023; Rogers, 2024).

O volume residual de respostas ausentes foi tratado por meio do FIML. O procedimento aproveita toda a informação observada de cada respondente e preserva os 621 casos da subamostra de calibração. A opção segue as diretrizes da literatura para tratamento de dados omissos sob o método de ML e suas variantes robustas (Brown, 2023; Kline, 2023).

A escala dos fatores latentes foi definida fixando-se suas variâncias em 1. Essa estratégia de identificação permite estimar livremente todas as cargas fatoriais e assegura a mesma métrica nos dois modelos alternativos. Ambas as especificações foram estimadas sobre a mesma subamostra de calibração e sob condições analíticas idênticas. A comparação empírica concentra-se, assim, nas três relações cruzadas adicionais postuladas no modelo H_1 (Jackson et al., 2009; Nye, 2022).

6.2 Parâmetros estimados

Antes da avaliação do ajuste, a Tabela 10 apresenta os parâmetros fatoriais estimados. Para cada relação entre fator e item, a tabela sintetiza a carga não padronizada, o erro-padrão robusto, a estatística z de teste e a carga padronizada, conforme recomendado por Jackson et al. (2009) e Nye (2022).

As cargas não padronizadas reportam a métrica original da escala. Por sua vez, as cargas padronizadas viabilizam a comparação direta da magnitude das relações em cada especificação, enquanto os erros-padrão asseguram inferência estatística corrigida para desvios de normalidade pelo método MLR.

A tabela consolida ambos os modelos, evidenciando as três relações adicionais de H_1 ao lado das cargas primárias. A avaliação dessas cargas conjuga os parâmetros estimados aos índices de ajuste global e local examinados nas seções subsequentes.

Tabela 10: Parâmetros fatoriais estimados nos modelos H_0 e H_1.

Modelo	Relação	Estimativa	EP	z	Padronizada
H0	psycho =~ Q5	0.523	0.033	15.774	0.639
H0	psycho =~ Q6	0.591	0.042	13.918	0.597
H0	psycho =~ Q7	0.561	0.031	18.139	0.653
H0	psycho =~ Q11	0.639	0.040	16.047	0.637
H0	psycho =~ Q19	0.764	0.033	23.296	0.798
H0	psycho =~ Q26	0.559	0.041	13.789	0.585
H0	physical =~ Q3	0.435	0.043	10.079	0.459
H0	physical =~ Q4	0.435	0.046	9.528	0.429
H0	physical =~ Q10	0.681	0.033	20.662	0.729
H0	physical =~ Q15	0.437	0.042	10.497	0.511
H0	physical =~ Q16	0.585	0.037	15.689	0.568
H0	physical =~ Q17	0.838	0.027	30.776	0.907
H0	physical =~ Q18	0.830	0.029	28.401	0.851
H0	social =~ Q20	0.711	0.033	21.654	0.806
H0	social =~ Q21	0.485	0.044	10.996	0.481

Tabela 10: Parâmetros fatoriais estimados nos modelos H_0 e H_1.

Modelo	Relação	Estimativa	EP	z	Padronizada
H0	social =~ Q22	0.637	0.035	18.180	0.727
H0	environment =~ Q8	0.448	0.041	10.919	0.455
H0	environment =~ Q9	0.453	0.037	12.206	0.529
H0	environment =~ Q12	0.751	0.041	18.249	0.688
H0	environment =~ Q13	0.508	0.038	13.259	0.624
H0	environment =~ Q14	0.631	0.039	16.006	0.667
H0	environment =~ Q23	0.525	0.041	12.939	0.535
H0	environment =~ Q24	0.455	0.042	10.814	0.478
H0	environment =~ Q25	0.491	0.043	11.407	0.504
H1	psycho =~ Q5	0.523	0.033	15.708	0.639
H1	psycho =~ Q6	0.593	0.043	13.923	0.600
H1	psycho =~ Q7	0.564	0.031	18.087	0.656
H1	psycho =~ Q11	0.640	0.040	16.068	0.639
H1	psycho =~ Q19	0.764	0.033	23.193	0.798
H1	psycho =~ Q26	0.559	0.041	13.779	0.585
H1	psycho =~ Q8	0.190	0.071	2.693	0.193
H1	psycho =~ Q9	0.126	0.056	2.235	0.147
H1	physical =~ Q3	0.427	0.043	9.991	0.451
H1	physical =~ Q4	0.430	0.046	9.432	0.424
H1	physical =~ Q10	0.676	0.033	20.293	0.722
H1	physical =~ Q15	0.253	0.049	5.206	0.296
H1	physical =~ Q16	0.584	0.037	15.638	0.568
H1	physical =~ Q17	0.846	0.027	31.669	0.915
H1	physical =~ Q18	0.834	0.029	29.055	0.855
H1	social =~ Q20	0.712	0.033	21.677	0.806
H1	social =~ Q21	0.485	0.044	10.998	0.480
H1	social =~ Q22	0.636	0.035	18.134	0.726
H1	environment =~ Q8	0.282	0.070	4.027	0.286
H1	environment =~ Q9	0.346	0.060	5.810	0.404
H1	environment =~ Q12	0.762	0.040	18.882	0.698
H1	environment =~ Q13	0.518	0.038	13.527	0.637
H1	environment =~ Q14	0.639	0.039	16.324	0.676
H1	environment =~ Q23	0.522	0.041	12.807	0.532
H1	environment =~ Q24	0.457	0.042	10.795	0.481
H1	environment =~ Q25	0.494	0.044	11.329	0.508
H1	environment =~ Q15	0.301	0.051	5.865	0.352

6.3 Convergência e admissibilidade

A verificação de convergência e admissibilidade encerra formalmente a etapa de estimação. A convergência atesta que o algoritmo encontrou uma solução estável para o modelo especificado. A admissibilidade matemática, por sua vez, é avaliada pela ausência de variâncias residuais negativas (casos Heywood) e de cargas padronizadas superiores à unidade (Brown, 2015; Nye, 2022). A Tabela 11 sintetiza essas verificações para H_0 e H_1, registrando o tamanho amostral efetivo e o número de parâmetros livres.

Na subamostra de calibração, ambos os modelos convergiram satisfatoriamente, sem apresentar estimativas impróprias de variância nem cargas padronizadas superiores a 1. As soluções mostram-se matematicamente admissíveis e aptas para as avaliações de ajuste global e local.

Tabela 11: Verificações de convergência e admissibilidade das soluções estimadas.

Modelo	N observado	Parâmetros livres	Convergência	Variâncias negativas	Cargas padronizadas > 1
H0	621	79	Sim	0	0
H1	621	82	Sim	0	0

7 Ajuste do modelo

7.1 Ajuste global dos modelos originais

O ajuste global avalia a extensão em que a estrutura de covariâncias dos indicadores é reproduzida pelas relações especificadas no modelo. Conforme fundamentado na Seção 2.5, essa avaliação articula o teste de ajuste exato a índices descritivos que quantificam a magnitude do desajuste residual.

A Tabela 12 apresenta os resultados de H_0 e H_1 estimados para os 621 participantes da subamostra de calibração. O teste χ^2 e o respectivo p-valor reportam a estatística com correção de escala (Satorra-Bentler), ao passo que CFI, TLI e RMSEA são informados em suas versões robustas, incluindo os limites do intervalo de confiança de 90% do RMSEA. O SRMR complementa o diagnóstico como medida de desajuste médio absoluto nas correlações. A leitura conjunta e integrada desses índices orienta o julgamento confirmatório das especificações originais (Mai et al., 2021; Niemand & Mai, 2018).

Tabela 12: Ajuste global dos modelos originais H_0 e H_1, estimados por MLR/FIML.

Medida	H0	H1
N	621	621
χ^2 ajust.	764.651	712.025
df	245	242
p-valor	< 0.001	< 0.001
CFI (robusto)	0.895	0.905
TLI (robusto)	0.882	0.892
RMSEA (robusto)	0.063	0.060
RMSEA IC90% inf.	0.058	0.055
RMSEA IC90% sup.	0.068	0.066
p (RMSEA $\leq$ 0,05)	< 0.001	< 0.001
SRMR	0.054	0.050

Os resultados da Tabela 12 evidenciam limitações de ajuste em ambas as especificações teóricas. O teste de ajuste exato revelou-se estatisticamente significante em H_0, $^2(245) =$ 764,651 (p <0,001), e em H_1, $^2(242) = 712{,}025$ (p <0,001). Como o teste χ^2 apresenta

alta sensibilidade ao tamanho amostral, rejeitando frequentemente discrepâncias de magnitude residual, sua interpretação é articulada aos índices descritivos (Nye, 2022; West et al., 2023).

Os índices incrementais concentram os principais indícios de inadequação nessa avaliação inicial. Em H_0, CFI e TLI robustos atingiram 0,895 e 0,882, ao passo que em H_1 registraram-se 0,905 e 0,892. Esses coeficientes atestam ganho modesto de ajuste frente ao modelo nulo de independência. Por sua vez, o RMSEA robusto situou-se em 0,063 (IC 90%: 0,058–0,068) em H_0 e em 0,060 (IC 90%: 0,055–0,066) em H_1, alinhando-se à zona de corte tradicional de 0,06.

O SRMR reportou comportamento mais favorável, marcando 0,054 em H_0 e 0,050 em H_1. A baixa magnitude média dos resíduos padronizados pode, no entanto, coexistir com focos de desajuste em pares específicos de itens, inspecionados oportunamente na etapa de diagnóstico local. A divergência entre as métricas reforça a necessidade metodológica de interpretá-las conjuntamente (Mai et al., 2021; Niemand & Mai, 2018).

H_1 apresenta índices descritivamente superiores, embora o ganho empírico frente a H_0 seja discreto. O déficit nos índices incrementais persiste nas duas especificações, nada obstante os níveis aceitáveis de RMSEA e SRMR. A abordagem por simulação complementa esse julgamento ao balizar as distribuições empíricas sob as condições da subamostra de calibração, conforme a estratégia discutida na Seção 2.5. O diagnóstico local examina os pontos em que as discrepâncias se concentram e norteia sua interpretação conceitual.

7.2 Confiabilidade dos escores

A confiabilidade mensura a precisão do escore compósito sob determinadas condições de mensuração. O coeficiente ômega (ω) figura como indicador central, estimado a partir das cargas fatoriais e das variâncias residuais do modelo. Essa aferição complementa o ajuste global e depende intrinsecamente da correta especificação fatorial (Cho, 2022; Flora, 2020).

A Tabela 13 expõe o ômega e o GLB para os quatro domínios em H_0 e H_1. Em H_1, o fator Psicológico incorpora também Q8 e Q9, enquanto o fator Meio Ambiente agrega Q15, passando a totalizar oito e nove itens, respectivamente. Os demais fatores preservam sua composição original. O GLB é computado separadamente com base na matriz observada de cada modelo, permitindo confrontar os resultados. Como sua dedução algébrica pressupõe independência residual e tende à superestimação amostral em amostras finitas, sua interpretação difere da confiabilidade derivada do modelo estrutural com covariância residual entre Q3 e Q4 (Trizano-Hermosilla & Alvarado, 2016).

Tabela 13: Confiabilidade dos domínios conforme a composição em H_0 e H_1

Modelo	Fator	Itens	Ômega	GLB
H0	psycho	6	0.807	0.844
H0	physical	7	0.800	0.883
H0	social	3	0.707	0.700
H0	environment	8	0.781	0.836
H1	psycho	8	0.643	0.858
H1	physical	7	0.729	0.883

Tabela 13: Confiabilidade dos domínios conforme a composição em H_0 e H_1

Modelo	Fator	Itens	Ômega	GLB
H1	social	3	0.707	0.700
H1	environment	9	0.673	0.851

No modelo de referência H_0, os coeficientes ômega dos quatro domínios situaram-se próximos ou acima do patamar aceitável de 0,70, registrando-se os menores índices no domínio Relações Sociais. Em H_1, a consistência dos escores Psicológico e Meio Ambiente recuou para níveis inferiores a essa referência. A avaliação desse recuo deve ponderar a incorporação de itens com cargas secundárias de baixa magnitude, somada à alteração na dimensionalidade teórica dos fatores. A confiabilidade complementa a caracterização psicométrica do instrumento, sem resolver as fragilidades de ajuste global verificadas anteriormente.

7.3 Distribuições dos índices sob os modelos empíricos

A análise de poder (*a priori*) baseou-se em parâmetros populacionais derivados da literatura e em pressupostos teóricos de planejamento. Uma vez realizada a estimação, as magnitudes de cargas e correlações empíricas afastaram-se daquelas premissas ideais. A avaliação por simulação de Monte Carlo foi, desse modo, reparametrizada pelas matrizes empíricas observadas, preservando a estrutura dos modelos e incorporando o tamanho amostral e as assimetrias distributivas da presente análise (Leite et al., 2023; Moshagen & Bader, 2024). Recalculam-se, assim, os limites fixados no planejamento para os 621 participantes da calibração, considerando os parâmetros estimados, as distribuições marginais e o padrão de dados omissos dessa subamostra. Esse procedimento desloca a questão do dimensionamento amostral prévio para o comportamento empírico esperado dos índices de ajuste sob as condições concretas do estudo.

Os parâmetros estimados delimitam as populações geradoras hipotéticas. Sob a primeira condição, os dados são gerados e modelados sob H_0, estabelecendo a distribuição empírica de referência dos índices quando as restrições teóricas do modelo são estritamente satisfeitas. Sob a segunda, os dados são gerados a partir de H_1 e estimados por H_0, avaliando-se a sensibilidade dos índices à omissão deliberada das três cargas cruzadas especificadas. Esse contraste confere continuidade ao protocolo de planejamento, atualizando as grandezas paramétricas à luz das estimativas da amostra.

A estratégia reúne a localização dos índices observados de H_0 frente à sua distribuição nula e a sensibilidade estatística dessa referência diante da má especificação concorrente. Os pontos de corte derivados dos quantis sob H_0 refletem a amostragem sob hipótese de ajuste exato, ao passo que a condição alternativa permite averiguar o poder de detecção da omissão estrutural postulada. O procedimento testa uma má especificação direcionada (as três cargas cruzadas), ao passo que a abordagem 3DFI exposta na Seção 2.5 introduz perturbações estocásticas distribuídas na matriz de correlações com magnitude calibrada por MAD (McNeish & Wolf, 2024). As conclusões mantêm-se condicionadas às estimativas pontuais e às distribuições simuladas. Tais estimativas permanecem fixas entre as replicações, cabendo ao Monte Carlo reproduzir o erro de amostragem condicional à subamostra de calibração. O diagnóstico local complementa essa análise ao rastrear anomalias pontuais que escapam ao escopo da alternativa geradora

modelada.

Processaram-se 1.000 replicações de 621 casos em cada condição geradora. A simulação replica as assimetrias e curtoses marginais observadas empiricamente, além de reproduzir o mapa exato de células ausentes da calibração sob o algoritmo MLR/FIML. A Figura 10 sintetiza as distribuições de amostragem para RMSEA, CFI e TLI robustos e SRMR, acompanhando as métricas reportadas na Tabela 12. A concentração de CFI e TLI em torno de 1 constitui o padrão esperado sob congruência perfeita entre gerador e estimador, permitindo mensurar objetivamente a distância em relação aos dados empíricos observados.

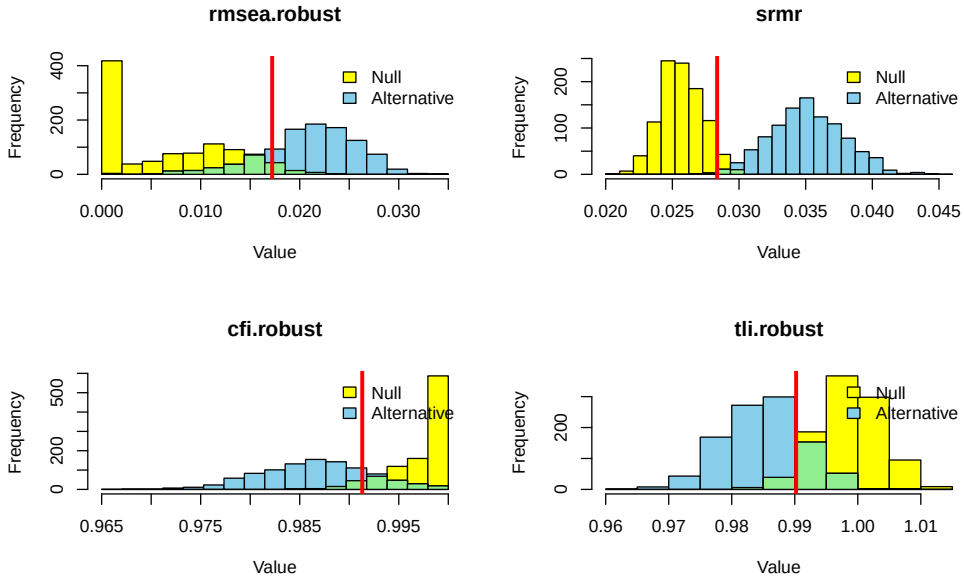


Figura 10: Distribuições de RMSEA, CFI e TLI robustos e SRMR em 1.000 replicações na amostra de calibração: geração sob H ou H e ajuste de H nas duas condições. H inclui as três cargas cruzadas omitidas pelo modelo de análise. As linhas vermelhas indicam os cortes sob H com = 0,05: percentil 95 para RMSEA/SRMR e percentil 5 para CFI/TLI.

A Figura 10 possibilita contrastar as distribuições teóricas sob H_0 com aquelas geradas sob a presença de três cargas cruzadas na população não especificadas no modelo analítico. Adotando-se o nível nominal de significância $\alpha = 0,05$, os limiares superiores de rejeição situam-se em 0,017 para o RMSEA e 0,028 para o SRMR. Paralelamente, os limites inferiores de corte equivalem a 0,991 para o CFI e 0,990 para o TLI. Ao incidir sobre a distribuição alternativa, esses cortes revelam coeficientes de poder estatístico correspondentes a 80,7%, 99,6%, 80,1% e 80,1%, respectivamente, mensurando a probabilidade do teste em detectar essa má especificação dimensional sob as condições amostrais do estudo.

Nos dados empíricos observados, H_0 reportou RMSEA de 0,063 e SRMR de 0,054, enquanto CFI e TLI atingiram 0,895 e 0,882. A projeção desses valores contra as curvas

simuladas atesta o afastamento substancial dos índices em relação aos limiares de ajuste exato. No entanto, a detecção da má especificação alternativa não elucida, isoladamente, a origem exata do desajuste observado nem comprova que esse decorra com exclusividade das três cargas cruzadas simuladas. Esses achados tornam obrigatório o aprofundamento do diagnóstico local, articulando a inspeção dos resíduos às premissas da teoria de mensuração e aos antecedentes empíricos da WHOQOL-Bref.

7.4 Diagnóstico local

O diagnóstico local tem por objetivo mapear as fontes específicas de discrepâncias empíricas que podem culminar no desajuste global observado. Essa investigação conjuga as estimativas fatoriais da Tabela 10 aos índices de modificação (MI), às mudanças esperadas nos parâmetros (EPC) e à matriz de resíduos de correlação, conforme prescrito na Seção 2.5. A análise concentra-se prioritariamente em H_0, modelo teórico de referência que orienta as decisões de reespecificação.

7.4.1 MI e EPC

O MI estima a redução teórica na estatística χ^2 proporcionada pela liberação individual de cada parâmetro previamente fixado. O EPC projeta a magnitude esperada da referida estimativa livre, e sua versão padronizada viabiliza a comparação direta do impacto paramétrico. Essas métricas devem ser interpretadas conjuntamente com as cargas fatoriais, com a magnitude dos resíduos e com o fundamento conceitual da relação analisada (Nye, 2022). A Tabela 14 consolida os vinte maiores MI de H_0 com seus respectivos EPC brutos e padronizados.

Tabela 14: Vinte maiores MI de H_0, com EPC bruto e padronizado.

Relação	MI	EPC	EPC padronizado
Q17 ~~ Q18	82.261	0.158	0.794
Q5 ~~ Q14	62.470	0.159	0.358
environment =~ Q5	48.605	0.362	0.442
environment =~ Q15	48.113	0.313	0.366
psycho =~ Q10	41.600	0.552	0.591
Q24 ~~ Q25	34.822	0.179	0.256
Q3 ~~ Q15	30.427	0.125	0.202
psycho =~ Q17	26.402	-0.408	-0.442
Q10 ~~ Q18	26.359	-0.088	-0.269
Q5 ~~ Q6	25.744	0.111	0.223
psycho =~ Q24	25.655	-0.324	-0.341
Q15 ~~ Q17	23.512	-0.076	-0.267
Q23 ~~ Q25	23.341	0.148	0.213
Q11 ~~ Q10	21.303	0.100	0.202
environment =~ Q17	20.282	-0.155	-0.167
physical =~ Q5	20.123	-0.377	-0.460
Q19 ~~ Q18	20.050	0.068	0.230
Q12 ~~ Q23	19.257	-0.139	-0.211
social =~ Q23	18.360	0.225	0.229
environment =~ Q10	18.223	0.177	0.189

7.4.2 Resíduos de correlação

Os resíduos de correlação indicam a magnitude em que cada associação bivariada entre indicadores é subestimada ou superestimada pela estrutura teórica do modelo. Essa métrica evidencia discrepâncias locais ocultas nos resumos globais (Crede & Harms, 2019; Nye, 2022).

A Figura 11 esquematiza a matriz de resíduos de H_0, dispondo os itens na sequência fatorial do modelo. A intensidade da gradação cromática traduz o valor absoluto do resíduo, ao passo que os valores numéricos informam o sinal aritmético da discrepância. Cada par bivariado é representado uma única vez. O limiar de relevância prática de $|r_{res}| \geq 0,10$, fundamentado na Seção 2.5, norteia o escrutínio conjunto com os MI, os EPC e a redação dos itens.

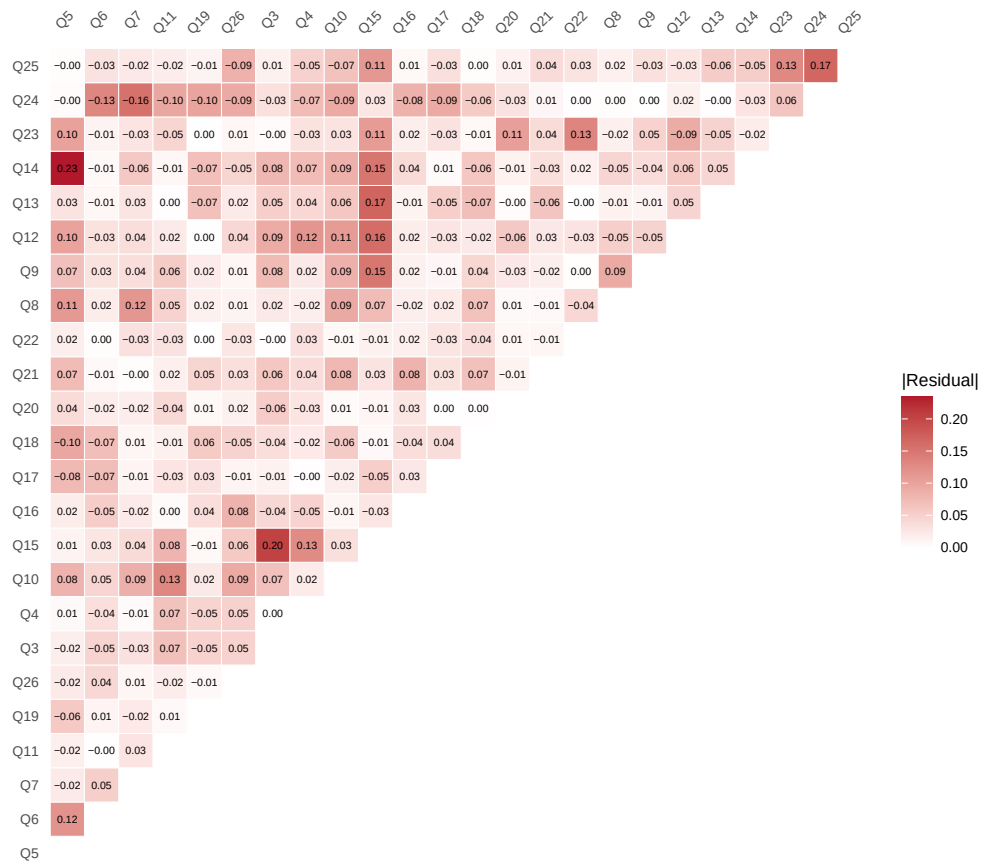


Figura 11: Resíduos de correlação de H . As cores representam magnitudes absolutas; os rótulos preservam o sinal e apresentam duas casas decimais.

O diagnóstico local de H_0 inspeciona inicialmente os três itens cujas cargas secundárias foram especificadas no modelo alternativo H_1. Em relação a Q15, a carga cruzada no fator Meio Ambiente figura com destaque entre os maiores MI (MI = 48,113, EPC padronizado = 0,366). O indicador apresenta também resíduos correlacionais positivos

expressivos com Q3 e Q4, da ordem de 0,20 e 0,13, indicando covariações bivariadas subestimadas pela estrutura confirmatória pura. Para os itens Q8 e Q9, sobressai a covariância residual entre seus termos de erro (MI = 10,847, EPC padronizado = 0,143), corroborada por resíduo de correlação positivo de 0,09. Esses dados expõem discrepâncias de mensuração nos itens englobados pela formulação de H_1, nada obstante a dependência residual entre Q8 e Q9 não comprovar, por si só, a existência de cargas adicionais no domínio Psicológico.

O escrutínio diagnóstico revela, ademais, pontos críticos de desajuste alheios às relações antecipadas por H_1. O item Q5 concentra indicações recorrentes de cargas secundárias nos domínios Físico e Meio Ambiente, além de correlação residual com Q6 e Q14. O resíduo de correlação no par Q5–Q14, de 0,235, estabelece a discrepância empírica de maior magnitude em toda a matriz. Por sua vez, o item Q24 combina carga primária modesta, demanda de carga cruzada no fator Psicológico e forte associação de erro com Q25, cujo resíduo de correlação alcança 0,166. Integrados aos desvios de Q15, esses achados qualificam Q5, Q15 e Q24 como os principais candidatos à investigação de modificações estruturais na calibração. O item Q8 mantém-se como candidato suplementar, dada a magnitude modesta de sua carga primária e a dependência residual com Q9. Paralelamente, as cargas atenuadas de Q3/Q4 e a covariação residual entre Q17 e Q18 também integram o inventário de desajustes, expressando particularidades de mensuração da escala.

A análise global evidenciou fragilidades de ajustamento nas especificações originais. O protocolo de simulação proveu distribuições de referência condicionais às propriedades da subamostra de calibração, balizando o julgamento dos índices obtidos. A leitura integrada de parâmetros, resíduos de correlação e indicadores de modificação isolou as fontes prioritárias de desajuste nos itens da escala. Essas evidências justificam a exploração de reespecificações, cujo critério de elegibilidade pressupõe a convergência com a teoria de mensuração e a integridade de cobertura dos domínios. Na Seção 8, essas hipóteses serão confrontadas com a literatura pregressa para fundamentar a especificação ajustada e suas diretrizes operacionais, sendo o modelo final validado de maneira independente na subamostra de validação.

8 Comparações e modificações dos modelos

8.1 Fundamentação das modificações

O diagnóstico da Seção 7, realizado na subamostra de calibração, destacou pontos de discrepância empírica que orientam modificações estruturais em H_0. A definição das alterações conjuga a magnitude das cargas primárias, a demanda empírica por cargas cruzadas, as correlações residuais entre termos de erro e o conteúdo conceitual dos indicadores, alinhando-se aos precedentes da literatura. A Tabela 15 sintetiza as variações de representação fatorial catalogadas por Lin & Yao (2022) e por Mosqueira Taipe et al. (2026). O levantamento abrange também indicadores que serão preservados, proporcionando um panorama das dificuldades psicométricas documentadas para o instrumento.

Tabela 15: Síntese das evidências na literatura sobre a representação dos indicadores do WHOQOL-Bref.

Prioridade de atenção na literatura	Item ou relação	Síntese das evidências na literatura
Alta	Q8 — segurança na vida diária	Apresenta ambiguidade entre os domínios ambiental e psicológico, com registros de cargas cruzadas, realocações e exclusões. É uma das divergências de alocação mais claras nas fontes consultadas.
Alta	Q10 — energia e fadiga	Há evidências de sobreposição entre os domínios físico e psicológico, expressas por cargas cruzadas e alocações alternativas.
Alta	Q14 — lazer	Apresenta registros de carga cruzada, realocação, exclusão e dependência residual com Q5, embora também tenha boa representação ambiental na síntese exploratória.
Alta	Q15 — mobilidade	Há registros de exclusão, inclusive em amostras de língua portuguesa, e de alocação em agrupamentos ambientais. Os achados indicam variação na representação da mobilidade entre contextos.
Moderada	Q3–Q4 — dor e necessidade de tratamento	Apresentam correlação residual recorrente e, em algumas soluções, formam uma dimensão própria. Sua retenção ou exclusão varia com o contexto clínico da amostra.
Moderada	Q5 — prazer de viver	Há registros de correlação residual com Q6 e Q14, alocação alternativa e exclusão em contextos específicos. A síntese exploratória também mostra forte representação do domínio psicológico.
Moderada	Q9 — ambiente físico	Apresenta registros de carga cruzada, correlação residual com Q8 e exclusão em contexto específico, indicando dificuldades na delimitação entre ambiente e percepção psicológica.
Moderada	Q19 — autoestima	Apresenta registros de correlação residual e alocações alternativas, ao lado de forte representação do domínio psicológico na síntese exploratória.
Moderada	Q26 — sentimentos negativos	Há registros de exclusão em diferentes contextos, embora a síntese exploratória também sustente sua representação no domínio psicológico.
Moderada	Q17–Q18 — atividades diárias e capacidade de trabalho	Há precedentes de covariância residual, cargas cruzadas e agrupamento em outra dimensão. A proximidade funcional e a dependência local constituem o argumento mais direto para a relação entre os itens.
Moderada	Q24–Q25 — acesso aos serviços e transporte	Há precedentes de correlação residual e de exclusão em uma solução reduzida, além de evidências de representação de ambos no domínio ambiental.

Tabela 15: Síntese das evidências na literatura sobre a representação dos indicadores do WHOQOL-Bref.

Prioridade de atenção na literatura	Item ou relação	Síntese das evidências na literatura
Moderada	Q12–Q13 — recursos financeiros e acesso à informação	Há precedentes de covariância residual entre os itens e de exclusão de ambos em uma solução reduzida. Os achados indicam variação na representação desses conteúdos ambientais.
Pontual	Q16 — sono e repouso	Há registro de exclusão em contexto específico. A evidência reunida oferece menor convergência para caracterizar uma dificuldade geral de representação.
Pontual	Q22 — apoio social	Há registros de exclusão em contextos específicos, mas a evidência reunida é menos consistente para caracterizar uma dificuldade geral de representação.

Fontes: Lin & Yao (2022) e Mosqueira Taipe et al. (2026). Os estudos individuais são conhecidos por intermédio dessas sínteses.

Nota: A prioridade de atenção na literatura considera a recorrência dos registros e suas implicações para a interpretação do modelo, atribuindo maior peso às cargas cruzadas, seguido da exclusão de itens (cargas baixas) e das covariâncias residuais. A classificação avalia o conjunto das evidências de cada foco. A seleção reúne achados destacados pelas revisões e precedentes recorrentes ou considerados relevantes.

A especificação modificada proposta para as análises subsequentes exclui os itens Q5, Q15 e Q24. Em Q5, a concentração de correlações residuais com Q6 e Q14 e as demandas de cargas cruzadas revelam dificuldades de delimitação dimensional. A literatura reporta essas dependências e alterações de alocação, ao passo que a síntese de Lin & Yao (2022) indica expressiva sobreposição com o domínio Psicológico. Sua exclusão responde diretamente às distorções locais detectadas na calibração. Quanto a Q15, a carga primária atenuada e a atração pelo domínio Meio Ambiente convergem com precedentes de eliminação em amostras lusófonas e propostas de alocação alternativa. O item Q24 conjuga carga reduzida e forte associação de erro com Q25, padrão igualmente documentado nos estudos prévios (Lin & Yao, 2022; Mosqueira Taipe et al., 2026).

Os itens Q3, Q4 e Q8 foram mantidos, nada obstante suas cargas modestas. Dor física e necessidade de tratamento médico representam facetas da saúde física cuja variância depende intrinsecamente do perfil clínico da amostra. A literatura relata a retenção de Q3 e Q4 em populações clínicas e sua eventual exclusão apenas em grupos jovens saudáveis. Preservam-se esses conteúdos conceituais e a covariância residual Q3–Q4 já contemplada na especificação original. Por sua vez, o item Q8 assegura a representação da segurança física no domínio Meio Ambiente. Sua proximidade empírica com o domínio Psicológico, apontada nas revisões, será considerada na interpretação do ajuste remanescente (Lin & Yao, 2022; Mosqueira Taipe et al., 2026).

Adicionou-se a covariância residual entre Q17 e Q18. A capacidade para o desempenho das atividades diárias e a aptidão laboral compartilham variância funcional específica, para além da dimensão física geral. A literatura documenta essa dependência residual em amostras de pacientes crônicos e de profissionais de saúde, sendo que o agrupamento conjunto desses indicadores em fatores funcionais específicos oferece evidência complementar dessa contiguidade. A inclusão desse parâmetro livre absorve a associação específica e viabiliza a permanência de ambos os itens no fator Físico (Lin & Yao, 2022; Mosqueira Taipe et al., 2026).

O modelo ajustado preserva quatro fatores correlacionados e os três indicadores sociais, passando a integrar 21 itens: 5 psicológicos, 6 físicos, 3 sociais e 7 ambientais. As exclusões suprimem a aferição direta de sentimentos positivos, locomoção e acesso aos serviços de saúde, devendo essa limitação acompanhar a interpretação dos escores. A formulação apoia-se nos diagnósticos da calibração e assume caráter exploratório. A especificação resultante é submetida ao teste confirmatório na subamostra de validação, avaliando-se o ajuste global, os desajustes locais e a confiabilidade compósita. A Figura 12 esquematiza a estrutura resultante com as covariâncias residuais especificadas.

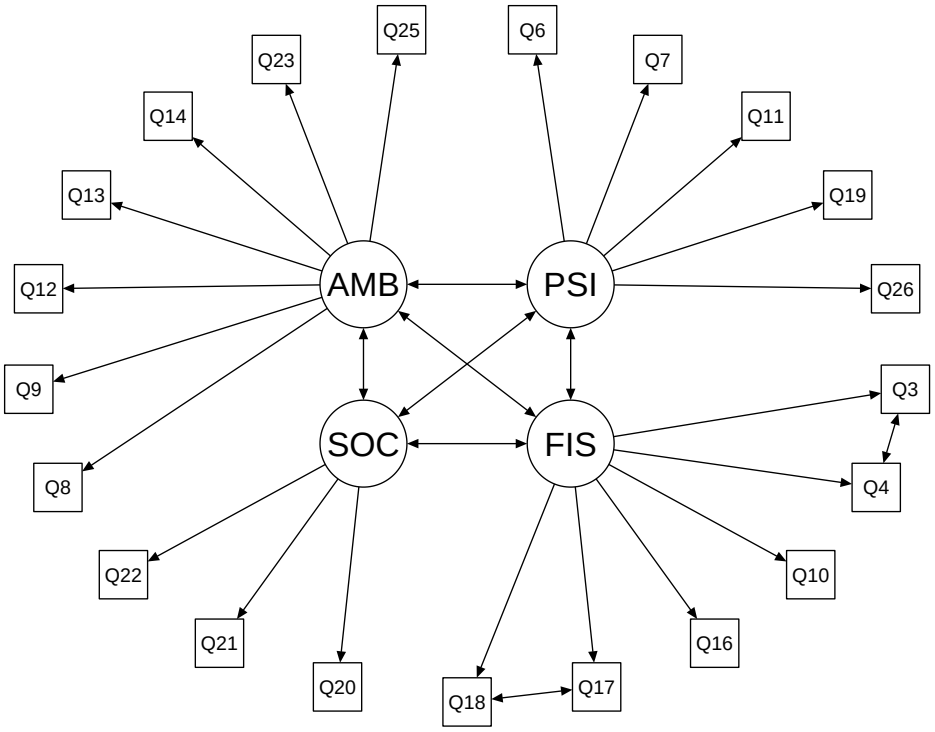


Figura 12: Modelo modificado de quatro fatores correlacionados e 21 indicadores, com covariâncias residuais Q3–Q4 e Q17–Q18. *Nota.* PSI = psicológico; FIS = físico; SOC = relações sociais; AMB = ambiente. Q5, Q15 e Q24 foram excluídos.

8.2 Parâmetros estimados

O modelo modificado é estimado sobre os 622 participantes da subamostra de validação, mantendo idênticas as condições analíticas delineadas na Seção 6. A avaliação reporta os parâmetros livres, verifica a admissibilidade matemática e examina o ajuste global e local da especificação formulada na calibração. A Tabela 16 sintetiza os parâmetros do modelo modificado.

Tabela 16: Parâmetros fatoriais estimados no modelo modificado.

Relação	Estimativa	EP	z	Padronizada
psycho =~ Q6	0.628	0.037	17.171	0.656
psycho =~ Q7	0.556	0.034	16.532	0.649
psycho =~ Q11	0.594	0.039	15.167	0.601
psycho =~ Q19	0.777	0.027	28.306	0.843
psycho =~ Q26	0.533	0.039	13.489	0.590
physical =~ Q3	0.350	0.040	8.734	0.375
physical =~ Q4	0.348	0.045	7.719	0.347
physical =~ Q10	0.670	0.030	22.676	0.772
physical =~ Q16	0.551	0.041	13.437	0.530
physical =~ Q17	0.711	0.033	21.532	0.775
physical =~ Q18	0.669	0.038	17.689	0.716
social =~ Q20	0.707	0.033	21.174	0.787
social =~ Q21	0.539	0.041	13.059	0.535
social =~ Q22	0.637	0.036	17.779	0.716
environment =~ Q8	0.490	0.043	11.264	0.504
environment =~ Q9	0.563	0.040	14.045	0.619
environment =~ Q12	0.755	0.043	17.630	0.683
environment =~ Q13	0.512	0.040	12.913	0.609
environment =~ Q14	0.667	0.039	17.259	0.692
environment =~ Q23	0.433	0.041	10.515	0.477
environment =~ Q25	0.463	0.044	10.557	0.472

As cargas fatoriais de Q3, Q4 e Q8 mantiveram-se modestas na validação: 0,375, 0,347 e 0,504, respectivamente. A decisão teórica de reter esses indicadores preserva os aspectos de dor, dependência terapêutica e segurança física. As magnitudes estimadas confirmam que tais facetas são capturadas apenas parcialmente pela estrutura latente comum. A validação sustenta a especificação fechada na calibração, englobando a covariância residual Q3–Q4 e parametrizando com transparência suas limitações operacionais.

O algoritmo convergiu plenamente para uma solução admissível, sem ocorrência de variâncias residuais negativas nem cargas padronizadas superiores à unidade. Esse diagnóstico atesta a regularidade matemática do ajuste, autorizando o avanço para a análise de ajuste global e local (Brown, 2015; Nye, 2022).

8.3 Ajuste global do modelo modificado

A Tabela 17 reporta os índices de ajuste global do modelo modificado: χ^2 com correção de escala, CFI, TLI e RMSEA robustos, intervalo de confiança de 90% do RMSEA e coeficiente SRMR, parametrizados para os 21 indicadores retidos.

Tabela 17: Ajuste global do modelo modificado, estimado por MLR/FIML.

Medida	Valor
N	622
χ^2 ajust.	437.311
df	181
p-valor	< 0.001
CFI (robusto)	0.938
TLI (robusto)	0.928
RMSEA (robusto)	0.052
RMSEA IC90% inf.	0.046
RMSEA IC90% sup.	0.058
p (RMSEA $\leq$ 0,05)	0.294
SRMR	0.041

A especificação ajustada exibiu CFI robusto de 0,938, TLI robusto de 0,928, RMSEA robusto de 0,052 e SRMR de 0,041. As magnitudes mostram-se expressivamente mais favoráveis em comparação às estimativas das estruturas originais detalhadas na Seção 7. Como esse confronto engloba composições de indicadores e subamostras distintas, ele não configura teste formal de superioridade paramétrica. O foco analítico repousa na adequação confirmatória da especificação modificada perante os dados de validação. No tocante ao ajuste estreito, o teste da hipótese nula de RMSEA robusto $\leq 0,05$ revelou $p = 0,294$, não se rejeitando a hipótese de ajuste próximo na subamostra de validação.

8.4 Confiabilidade dos escores

A Tabela 18 expõe os coeficientes ômega e GLB calculados para os quatro escores com os 21 itens retidos. O ômega reproduz a parametrização do modelo modificado, incorporando as covariâncias residuais Q3–Q4 e Q17–Q18, ao passo que o GLB repousa na premissa algébrica de independência residual (Cho, 2022; Flora, 2020; Trizano-Hermosilla & Alvarado, 2016).

Tabela 18: Confiabilidade dos escores no modelo modificado.

Fator	Itens	Ômega	GLB
psycho	5	0.798	0.823
physical	6	0.694	0.854
social	3	0.717	0.712
environment	7	0.780	0.838

A reespecificação redefiniu a composição dos escores Psicológico, Físico e Meio Ambiente, mantendo-se intacto o fator Relações Sociais com seus três itens. Na subamostra de validação, os coeficientes ômega dos escores Psicológico, Social e Meio Ambiente suplantaram o limiar de 0,70. No domínio Físico, a estimativa fixou-se em 0,694, situando-se marginalmente abaixo dessa referência. Os índices GLB superaram 0,70 em todos os quatro domínios (Cho, 2022; Trizano-Hermosilla & Alvarado, 2016).

As estimativas de ômega oscilaram entre 0,694 e 0,798, enquanto os valores de GLB situaram-se entre 0,712 e 0,854. A interpretação desses números perante a estrutura original H_0 deve considerar o estreitamento conceitual decorrente da exclusão de itens e a independência da amostra de teste.

8.5 Avaliação por 3DFI

A avaliação confirmatória da especificação ajustada na subamostra de validação fundamenta-se na metodologia 3DFI exposta na Seção 2.5. A calibração de pontos de corte dinâmicos qualifica o grau de ajuste aproximado mediante a indução controlada de discrepâncias médias absolutas nas correlações, sem necessidade de arbitrar trajetórias específicas de má especificação. Ao passo que a simulação descrita na Seção 7 confrontou o modelo com uma alternativa estrutural pré-determinada, o objetivo da presente análise consiste em mensurar a magnitude tolerável do desajuste residual remanescente (McNeish & Wolf, 2024). Trata-se de uma avaliação graduada da solução após o refinamento exploratório, respondendo a perguntas complementares: a distância em relação ao ajuste exato nas estruturas iniciais e o nível de concordância empírica aproximada no modelo modificado.

O procedimento simula 1.000 amostras com base nos 21 indicadores, incorporando as assimetrias distributivas empíricas sob o estimador MLR. Fixaram-se três níveis médios de discrepância absoluta (MAD): 0,038, 0,050 e 0,060, representativos dos patamares de ajuste estreito (*close*), aceitável (*fair*) e medíocre (*mediocre*), segundo as diretrizes de McNeish & Wolf (2024). A Tabela 19 consolida os pontos de corte dinâmicos para CFI, RMSEA e limite superior do intervalo de confiança de 90% do RMSEA.

Tabela 19: Limites 3DFI calibrados para o modelo modificado.

Referência	MAD	MAD simulado	CFI	RMSEA	RMSEA IC90% sup.
Próximo	0.038	0.036	0.975	0.032	0.040
Razoável	0.050	0.048	0.924	0.058	0.064
Medíocre	0.060	0.058	0.875	0.076	0.082

A Figura 13 ilustra a referência de ajuste aceitável (*fair*), confrontando as distribuições simuladas de CFI e RMSEA na condição nula e sob a perturbação programada. O contraste entre limites dinâmicos e referências convencionais demonstra a sensibilidade das métricas às propriedades empíricas da amostra.

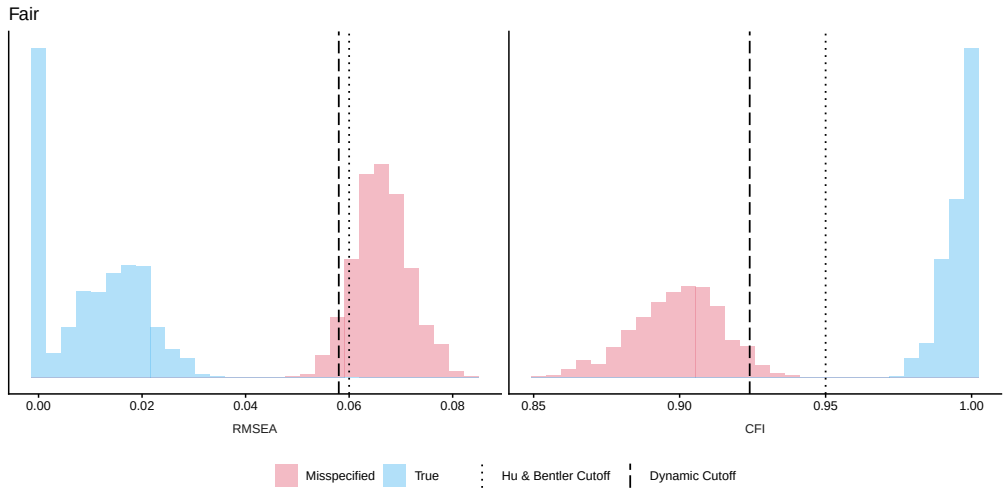


Figura 13: Distribuições simuladas de CFI e RMSEA e limites dinâmicos da referência razoável (fair), com limites convencionais para comparação.

Os coeficientes empíricos observados cumprem com folga os critérios da referência aceitável (*fair*): CFI de pelo menos 0,924, RMSEA de até 0,058 e limite superior do intervalo de confiança de 90% do RMSEA de até 0,064. A referência de ajuste estreito (*close*), mais rigorosa, não foi plenamente atingida. Desse modo, o procedimento 3DFI classifica o comportamento do modelo modificado como aceitável perante os dados de validação.

8.6 Diagnóstico local

O diagnóstico local mapeia os desajustes residuais remanescentes no modelo modificado sob a amostra de validação. A Tabela 20 elenca os vinte maiores MI acompanhados de seus respectivos EPC brutos e padronizados. Esses indicadores sinalizam o ganho potencial derivado da liberação isolada de parâmetros e devem ser cotejados com as cargas fatoriais e a matriz de resíduos (Crede & Harms, 2019; Nye, 2022).

Tabela 20: Vinte maiores índices de modificação do modelo modificado, com EPC bruto e padronizado.

Relação	MI	EPC	EPC padronizado
Q8 ~~ Q9	38.123	0.169	0.282
Q9 ~~ Q23	19.371	0.114	0.200
Q17 ~~ Q22	18.096	-0.062	-0.171
physical =~ Q6	17.080	-1.197	-1.251
Q12 ~~ Q14	17.060	0.126	0.225
Q20 ~~ Q12	14.926	-0.093	-0.208
Q19 ~~ Q10	14.908	-0.062	-0.225
social =~ Q6	13.327	0.234	0.245
environment =~ Q10	12.885	0.184	0.212
psycho =~ Q18	12.786	0.825	0.884

Tabela 20: Vinte maiores índices de modificação do modelo modificado, com EPC bruto e padronizado.

Relação	MI	EPC	EPC padronizado
Q9 ~~ Q12	12.119	-0.102	-0.176
Q18 ~~ Q22	12.003	0.055	0.135
Q19 ~~ Q3	11.631	-0.063	-0.147
Q23 ~~ Q25	11.594	0.102	0.147
Q8 ~~ Q13	11.534	-0.086	-0.154
Q6 ~~ Q3	11.394	-0.081	-0.130
social =~ Q26	11.212	-0.211	-0.233
Q12 ~~ Q13	10.791	0.089	0.165
social =~ Q7	9.959	-0.182	-0.212
Q13 ~~ Q14	9.786	0.074	0.159

A Figura 14 exibe as divergências entre as correlações bivariadas observadas e aquelas reproduzidas pela estrutura de 21 itens. Cada par é representado uma única vez, associando-se a saturação de cor à magnitude absoluta do resíduo. O limiar prático de $|r_{res}| \geq 0,10$ serve de guia para identificação dos pontos residuais de desajuste.

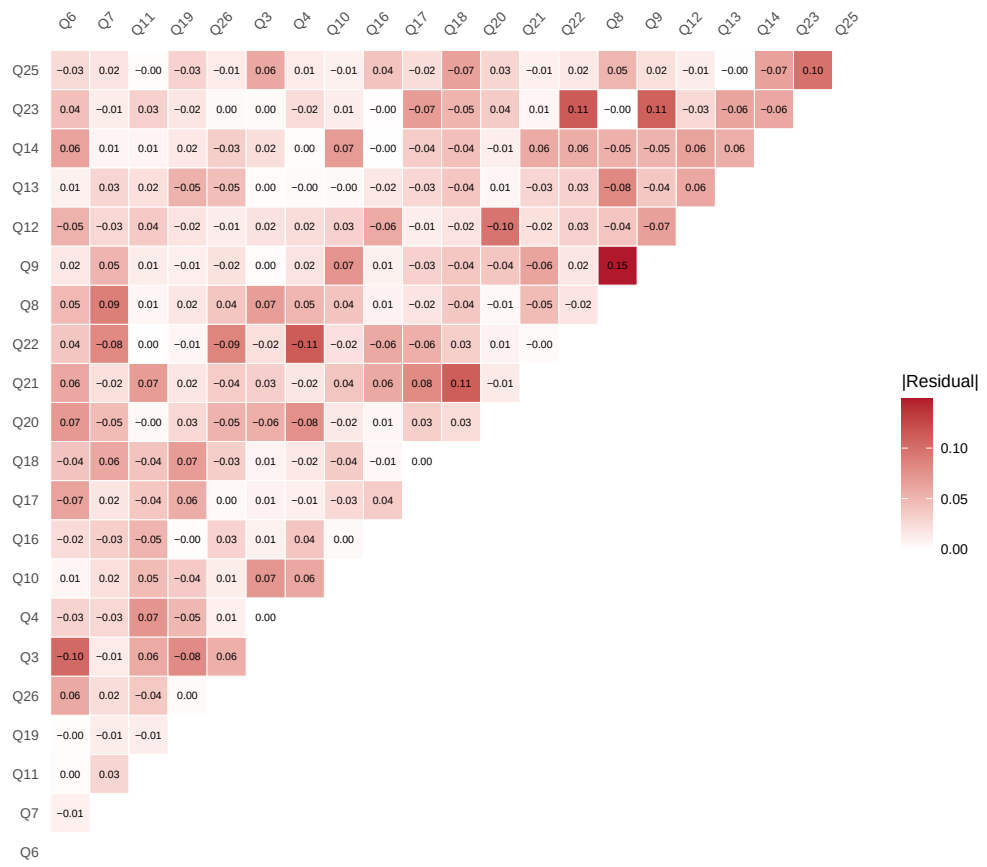


Figura 14: Resíduos de correlação do modelo modificado. As cores representam magnitudes absolutas; os rótulos preservam o sinal.

O resíduo absoluto de maior magnitude fixou-se em 0,149, registrando-se que apenas 6 dos 210 pares avaliados ultrapassam o limiar prático de 0,10. A integração desse mapa com os parâmetros fatoriais e a tabela de MI/EPC delimita com clareza as discrepâncias remanescentes. O par Q8–Q9 lidera tanto o ranking de MI quanto a maior discrepância correlacional isolada, acompanhado de perto pela associação entre Q9 e Q23. A demanda de carga cruzada do item Q6 no fator Físico atinge EPC padronizado de -1,251, figurando entre as ressalvas diagnósticas documentadas. Cumpre salientar que esses índices projetam estimativas pontuais sob liberação univariada, não representando parâmetros de um modelo reajustado.

A reespecificação consolidou um modelo parcimonioso de quatro fatores com 21 indicadores, resguardando os três itens do domínio Social, mantendo a covariância de erro Q3–Q4 e incorporando a dependência entre Q17 e Q18. Na subamostra de validação, a solução satisfez os critérios de ajuste aceitável do 3DFI e não teve rejeitada a hipótese de ajuste próximo pelo teste do RMSEA. Os coeficientes ômega superaram 0,70 nos domínios Psicológico, Social e Meio Ambiente. No domínio Físico, o ômega ficou abaixo dessa referência, conforme reportado na análise de confiabilidade. Adota-se essa especificação, fundamentada na calibração e avaliada na validação, para orientar as conclusões meto-

dológicas do estudo, recomendando-se cautela quanto à cobertura conceitual reduzida dos domínios.

8.7 Invariância de medida

A etapa conclusiva afere a estabilidade estrutural do modelo de 21 indicadores ao longo da partição amostral entre calibração e validação. A análise fatorial confirmatória multigrupo assegura a identificação simultânea de ambas as subamostras, examinando a comparabilidade dos parâmetros de mensuração. O nível configural afere se a configuração de fatores e indicadores mantém-se invariante entre os grupos, enquanto o nível métrico impõe restrições de igualdade sobre as cargas fatoriais não padronizadas (Fischer & Karl, 2019; Putnick & Bornstein, 2016).

Embora a invariância métrica seja condição suficiente para a estabilidade estrutural das cargas fatoriais, os níveis escalar e estrito estendem a avaliação, impondo sucessivamente a equivalência dos interceptos e das variâncias residuais (Putnick & Bornstein, 2016). Todas as especificações mantêm o estimador MLR/FIML e as covariâncias residuais Q3–Q4 e Q17–Q18 livres em cada subamostra.

A Tabela 21 apresenta os testes dos modelos e as diferenças robustas entre níveis hierárquicos. A avaliação considera o desajuste decorrente das restrições adicionais, distinguindo o teste de diferença da hipótese de ajuste exato de cada especificação sob correção robusta (Brown, 2015).

Tabela 21: Testes dos modelos de invariância e diferenças robustas entre níveis hierárquicos.

Modelo	χ^2 ajust.	gl	$\Delta\chi^2$ ajust.	Δgl	p da diferença
Configural	834.452	362	–	–	–
Métrico	847.998	379	12.521	17	0.768
Escalar	869.482	396	18.625	17	0.350
Estrito	895.285	417	26.812	21	0.177

A restrição de igualdade sobre as cargas fatoriais foi sustentada pelos dados: o contraste entre os modelos métrico e configural exibiu $\Delta\chi^2$ robusto de 12,521, com 17 graus de liberdade e $p = 0,768$. As etapas subsequentes mostraram-se igualmente consistentes: os contrastes escalar/métrico e estrito/escalar retornaram $p = 0,350$ e $p = 0,177$, respectivamente. Esse padrão empírico respalda a invariância métrica e estrita da versão reduzida do instrumento (Putnick & Bornstein, 2016).

A Tabela 22 consolida a trajetória dos índices de ajuste e suas respectivas variações incrementais (Δ).

Tabela 22: Índices de ajuste e suas mudanças entre níveis hierárquicos de invariância: CFI e RMSEA robustos e SRMR.

Modelo	CFI	ΔCFI	RM-SEA	ΔRMSEA	SRMR	ΔSRMR
Configural	0.943	–	0.050	–	0.040	–
Métrico	0.944	+0.0005	0.048	-0.0013	0.042	+0.0028
Escalar	0.943	-0.0002	0.047	-0.0010	0.043	+0.0005
Estrito	0.943	-0.0007	0.047	-0.0009	0.043	+0.0004

A especificação configural registrou CFI robusto de 0,943, RMSEA robusto de 0,050 e SRMR de 0,040. A fixação de igualdade sobre as cargas fatoriais aumentou o CFI em 0,0005 e atenuou o RMSEA em 0,0013, observando-se variação mínima no SRMR (0,0028). Esse comportamento indica que a imposição de invariância métrica não degradou a qualidade do ajuste.

Nos níveis escalar e estrito, os recuos observados no CFI fixaram-se em 0,0002 e 0,0007, valores inferiores ao limiar convencional de alerta de 0,010 (Chen, 2007; Cheung & Rensvold, 2002). O alinhamento entre os testes formais de diferença e a pequena magnitude das variações descritivas sustenta evidências de invariância de mensuração entre as amostras de calibração e validação (Putnick & Bornstein, 2016).

A equivalência métrica foi igualmente corroborada por teste de permutação estratificada para as variações de CFI, RMSEA e SRMR (Jorgensen et al., 2018). Para tanto, redistribuíram-se aleatoriamente os rótulos de calibração e validação dentro de cada estrato de integridade de dados (completos vs. omissos), reestimando-se os modelos configural e métrico para averiguar a probabilidade empírica de observar alterações adversas nos índices.

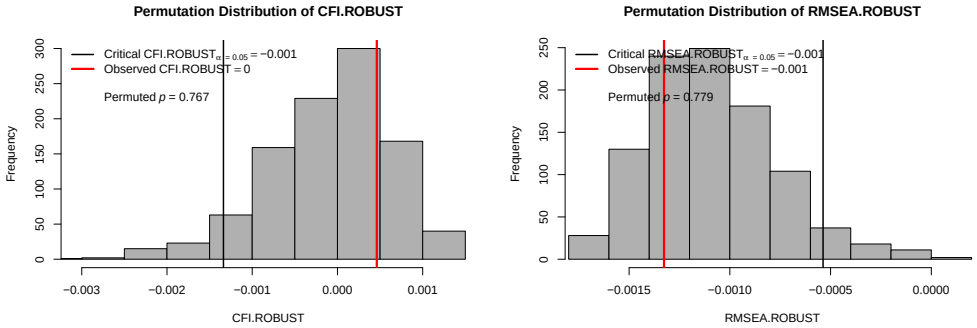
A Tabela 23 resume as variações observadas e os valores-p derivados das 1000 permutações empíricas, das quais 1000 alcançaram plena convergência numérica.

Tabela 23: Diferenças observadas e valores-p da permutação para a invariância métrica.

Índice	Δ observado	p da permutação
CFI (robusto)	+0.0005	0.767
RMSEA (robusto)	-0.0013	0.779
SRMR	+0.0028	0.559

Os valores-p empíricos resultaram em 0,767 para o CFI robusto, 0,779 para o RMSEA robusto e 0,559 para o SRMR. As alterações observadas situaram-se bem distantes das regiões de cauda associadas à degradação do ajuste, em concordância com os resultados dos testes de razão de verossimilhanças (Jorgensen et al., 2018).

A Figura 15 ilustra as distribuições empíricas de ΔCFI e ΔRMSEA. As linhas tracejadas vermelhas demarcam os valores observados na amostra, ao passo que as linhas pretas sinalizam os pontos de corte críticos a 5%. Em ambos os casos, as estimativas empíricas permanecem afastadas da zona de rejeição.



(a) CFI robusto

(b) RMSEA robusto

Figura 15: Distribuições de permutação das mudanças de CFI e RMSEA robustos na avaliação da invariância métrica.

A análise multigrupo conduzida entre as subamostras de calibração e validação sustentou evidências de invariância métrica da estrutura de 21 indicadores, com resultados igualmente favoráveis às restrições escalares e estritas. Esses resultados complementam as evidências de ajuste global, confiabilidade e diagnóstico local e indicam estabilidade interna entre parcelas da mesma base de dados. Seu alcance considera o histórico de seleção do modelo na amostra completa (Fischer & Karl, 2019; Putnick & Bornstein, 2016).

9 Conclusões

A condução da AFC exige articular as evidências prévias sobre o instrumento às decisões que sustentam seu emprego em modelos estruturais. Nesse artigo-tutorial, sistematizou-se esse percurso a partir de especificações do WHOQOL-Bref fundamentadas na literatura, conectando a seleção de uma estrutura populacional plausível e de suas más especificações ao planejamento amostral. O tratamento contínuo dos itens, fundamentado nas categorias de resposta e em suas propriedades psicométricas, permitiu integrar estimação robusta, tratamento de dados ausentes e procedimentos de avaliação do modelo. Assim, apresentou-se uma aplicação em que as escolhas analíticas permaneceram vinculadas aos objetivos da mensuração.

O planejamento amostral foi desenvolvido em uma sequência que partiu da referência analítica de poder, explorou uma faixa de tamanhos amostrais e concentrou as simulações em um tamanho fixo. Essa organização permitiu examinar a capacidade de detectar discrepâncias previamente especificadas sob condições de não normalidade e dados ausentes. Adicionalmente, incorporou-se a qualidade das respostas ao percurso analítico por meio da abordagem de C/IER, articulando procedimentos previstos no questionário a indicadores estatísticos organizados em obstáculos. Dessa forma, documentaram-se tanto as condições consideradas no planejamento quanto os critérios empregados para constituir a amostra analítica.

A avaliação do ajuste reuniu evidências globais, confiabilidade e diagnóstico local, complementadas por referências obtidas por simulação. As simulações empíricas permitiram contextualizar o comportamento dos índices nas condições examinadas, enquanto o

3DFI exemplificou a avaliação do modelo modificado por limites específicos de ajuste. As reespecificações decorreram do confronto entre os diagnósticos da calibração e os problemas de mensuração documentados na literatura, explicitando-se a justificativa das exclusões e das covariâncias residuais. Essa sequência tornou verificável a passagem entre a identificação do desajuste, a formulação das modificações e a avaliação posterior da estrutura retida.

Na amostra de validação, o modelo de quatro fatores e 21 indicadores apresentou evidências favoráveis de ajuste, e a análise multigrupo sustentou a estabilidade das cargas entre calibração e validação. Os resultados da permutação estratificada foram compatíveis com a invariância métrica, acompanhando as evidências dos testes de diferença e das mudanças nos índices. O alcance desses achados corresponde à estabilidade interna entre amostras da mesma base, com histórico de seleção do modelo na amostra completa.

A contribuição didática e computacional do tutorial reside na integração entre fundamentação metodológica, código executável e interpretação de cada resultado. No HTML, os blocos expansíveis permitem examinar a finalidade, a preparação e a execução das análises, enquanto o PDF preserva a continuidade da narrativa científica. A organização segundo o ARTE conecta dados brutos, funções reutilizáveis, resultados e documentos-fonte, com dependências e instruções de execução documentadas. Essa infraestrutura concretiza práticas de Ciência Aberta ao permitir o exame das decisões, a reprodução da aplicação e a adaptação dos materiais a outros problemas de mensuração.

O percurso apresentado oferece a pesquisadores e estudantes uma referência para planejar, executar e comunicar AFCs com justificativas explícitas e procedimentos reproduzíveis. Sua reutilização requer adequar os modelos, os critérios de qualidade das respostas e as condições de simulação às características de cada investigação. As rotinas modulares e a redação dinâmica em Quarto permitem acompanhar essas adaptações e manter a correspondência entre análises e resultados reportados. Dessa forma, o artigo pode fornecer uma estrutura de trabalho para reunir evidências sobre modelos de medida, e seu uso pode sustentar questões teóricas que orientam a pesquisa de uma gama ampla de cientistas sociais.

Referências

AERA, APA, & NCME. (1999). *Standards for Educational and Psychological Testing*. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.

AERA, APA, & NCME. (2014). *Standards for Educational and Pshychological Testing*. American Educational Research Association, American Psychological Association & National Council on Measurement in Education.

Arthur, W., Hagen, E., & George, F. (2021). The Lazy or Dishonest Respondent: Detection and Prevention. *Annu. Rev. Organ. Psychol. Organ. Behav.*, *8*, 105–137. https://doi.org/10.1146/annurev-orgpsych-012420-055324

Bandalos, D. L. (2014). Relative Performance of Categorical Diagonally Weighted Least Squares and Robust Maximum Likelihood Estimation. *Structural Equation Modeling*, *21*(1), 102–116. https://doi.org/10.1080/10705511.2014.859510

Bandalos, D. L. (2018). *Measurement Theory and Applications for the Social Sciences*. Guilford Press.

Bell, S. M., Chalmers, R. P., & Flora, D. B. (2023). The Impact of Measu-

rement Model Misspecification on Coefficient Omega Estimates of Composite Reliability. *Educational and Psychological Measurement*, 1–36. https://doi.org/10.1177/00131644231155804

Biemann, T., Koch-Bayram, I. F., Meier-Barthold, M., & Aguinis, H. (2025). Using Markov Chains to Detect Careless Responding in Survey Research. *Organ. Res. Methods*, *28*(4), 543–568. https://doi.org/10.1177/10944281251334778

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. The Guilford Press.

Brown, T. A. (2023). Confirmatory Factor Analysis. Em R. H. Hoyle (Org.), *Handbook of Structural Equation Modeling*. The Guilford Press.

Byrne, B. M. (2016). *Structural Equation Modeling with AMOS: Basic Concepts, Application, and Programming*. Routledge Taylor & Francis Group.

Chauliac, M., Willems, J., Gijbels, D., & Donche, V. (2023). The Prevalence of Careless Response Behaviour and Its Consequences on Data Quality in Self-Report Questionnaires on Student Learning. *Front. Educ.*, *8*. https://doi.org/10.3389/feduc.2023.1197324

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cho, E. (2022). Reliability and Omega Hierarchical in Multidimensional Data: A Comparison of Various Estimators. *Psychological Methods*. https://doi.org/10.1037/met0000525

Cohen, R. J., Schneider, J. W., & Tobin, R. M. (2022). *Psychological Testing and Assessment: An Introduction to Test and Measurement*. McGraw Hill LLC.

Collier, J. E. (2020). *Applied Structural Equation Modeling Using AMOS: Basic to Advanced Techniques*. Routledge.

Crede, M., & Harms, P. (2019). Questionable Research Practices When Using Confirmatory Factor Analysis. *Journal of Managerial Psychology*, *34*(1), 18–30. https://doi.org/10.1108/JMP-06-2018-0272

Curran, P. G. (2016). Methods for the Detection of Carelessly Invalid Responses in Survey Data. *J. Exp. Soc. Psychol.*, *66*, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006

Davvetas, V., Diamantopoulos, A., Zaefarian, G., & Sichtmann, C. (2020). Ten Basic Questions about Structural Equations Modeling You Should Know the Answers to – But Perhaps You Don't. *Industrial Marketing Management*, *90*, 252–263. https://doi.org/10.1016/j.indmarman.2020.07.016

DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best Practice Recommendations for Data Screening. *J. Organ. Behav.*, *36*(2), 171–181. https://doi.org/10.1002/job.1962

Dunn, A. M., Heggestad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-Individual Response Variability as an Indicator of Insufficient Effort Responding: Comparison to Other Indicators and Relationships with Individual Differences. *J. Bus. Psychol.*, *33*(1), 105–121. https://doi.org/10.1007/s10869-016-9479-0

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From Alpha to Omega: A Practical Solution to the Pervasive Problem of Internal Consistency Estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Enders, C. (2023). Fitting Structural Equation Models with Missing Data. Em R. H.

Hoyle (Org.), *Handbook of Structural Equation Modeling*. The Guilford Press.
Feng, Y., & Hancock, G. R. (2023). Power Analysis within a Structural Equation Modeling Framework. Em R. H. Hoyle (Org.), *Handbook of Structural Equation Modeling*. The Guilford Press.
Fischer, R., & Karl, J. A. (2019). A Primer to (Cross-Cultural) Multi-Group Invariance Testing Possibilities in R. *Frontiers in Psychology*, *10*, 1507. https://doi.org/10.3389/fpsyg.2019.01507
Flake, J. K., Davidson, I. J., Wong, O., & Pek, J. (2022). Construct Validity and the Validity of Replication Studies: A Systematic Review. *American Psychologist*, *77*(4), 576–588. https://doi.org/10.1037/amp0001006
Flake, J. K., & Fried, E. I. (2020). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393
Flake, J. K., Pek, J., & Hehman, E. (2017). Construct Validation in Social and Personality Research: Current Practices and Recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378. https://doi.org/10.1177/1948550617693063
Flora, D. B. (2020). Your Coefficient Alpha Is Probably Wrong, but Which Coefficient Omega Is Right? A Tutorial on Using R to Obtain Better Reliability Estimates. *Advances in Methods and Practices in Psychological Science*, *3*(4), 484–501. https://doi.org/10.1177/2515245920951747
Flora, D. B., & Flake, J. K. (2017). The Purpose and Practice of Exploratory and Confirmatory Factor Analysis in Psychological Research: Decisions for Scale Development and Validation. *Canadian Journal of Behavioural Science*, *49*(2), 78–88. https://doi.org/10.1037/cbs0000069
Flora, D. B., LaBrish, C., & Chalmers, R. P. (2012). Old and New Ideas for Data Screening and Assumption Testing for Exploratory and Confirmatory Factor Analysis. *Frontiers in Psychology*, *3*(MAR), 55. https://doi.org/10.3389/fpsyg.2012.00055
Furr, M. R. (2021). *Psychometrics: An Introduction*. SAGE Publications.
Gilroy, S. P., & Kaplan, B. A. (2019). Furthering Open Science in Behavior Analysis: An Introduction and Tutorial for Using GitHub in Research. *Perspectives on Behavior Science*, *42*(3), 565–581. https://doi.org/10.1007/s40614-019-00202-5
Goldammer, P., Annen, H., Stöckli, P. L., & Jonas, K. (2020). Careless Responding in Questionnaire Measures: Detection, Impact, and Remedies. *Leadersh. Q.*, *31*(4), 101384. https://doi.org/10.1016/j.leaqua.2020.101384
Goodboy, A. K., & Martin, M. M. (2020). Omega over Alpha for Reliability Estimation of Unidimensional Communication Measures. *Annals of the International Communication Association*, *44*(4), 422–439. https://doi.org/10.1080/23808985.2020.1846135
Green, S. B., & Yang, Y. (2015). Evaluation of Dimensionality in the Assessment of Internal Consistency Reliability: Coefficient Alpha and Omega Coefficients. *Educational Measurement: Issues and Practice*, *34*(4), 14–20. https://doi.org/10.1111/emip.12100
Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing. *Marketing Science*, *23*(4), 519–529. https://doi.org/10.1287/mksc.1040.0070
Groskurth, K., Bluemke, M., & Lechner, C. M. (2024). Why We Need to Abandon Fixed Cutoffs for Goodness-of-Fit Indices: An Extensive Simulation and Possible Solutions. *Behav. Res. Methods*, *56*(4), 3891–3914. https://doi.org/10.3758/s13428-023-02193-3
Hair, J. F., Hult, T. M. G., Ringle, C. M., & Sarstedt, M. (2022). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Sage Publications.
Hair, J. F., Sarstedt, M., Ringle, C., & Gudergan, S. P. (2017). *Advanced Issues in*

Partial Least Squares Structural Equation Modeling. SAGE Publications, Inc.
Hair, J., Black, W., Babin, B., & Anderson, R. (2019). *Multivariate Data Analysis*. Cengage Learning.
Harpe, S. E. (2015). How to Analyze Likert and Other Rating Scale Data. *Curr. Pharm. Teach. Learn.*, *7*(6), 836–850. https://doi.org/10.1016/j.cptl.2015.08.001
Harrington, D. (2009). *Confirmatory Factor Analysis*. Oxford University Press.
Hayes, A. F., & Coutts, J. J. (2020). Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But…. *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629
Henseler, J. (2021). *Composite-Based Structural Equation Modeling: Analyzing Latent and Emergent Variables*. The Guilford Press.
Holgado-Tello, F., Morata-Ramirez, M., & García, M. (2016). Robust Estimation Methods in Confirmatory Factor Analysis of Likert Scales: A Simulation Study. *International Review of Social Sciences and Humanities*, *11*(2), 80–96.
Hong, M., Steedle, J. T., & Cheng, Y. (2020). Methods of Detecting Insufficient Effort Responding: Comparisons and Practical Recommendations. *Educ. Psychol. Meas.*, *80*(2), 312–345. https://doi.org/10.1177/0013164419865316
Hoyle, R. H. (2023). Structural Equation Modeling: An Overview. Em R. H. Hoyle (Org.), *Handbook of Structural Equation Modeling*. Guilford Press.
Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient Effort Responding: Examining an Insidious Confound in Survey Data. *J. Appl. Psychol.*, *100*(3), 828–845. https://doi.org/http://dx.doi.org/10.1037/a0038510
Huang, J. L., Wang, Z., Huang, R., Wu, D., & Shi, H. (2026). *Insufficient Effort Responding in Management Research: A Critical Review and Future Directions*. *52*(1). https://doi.org/10.1177/01492063251330268
Huang, P.-H. (2017). Asymptotics of AIC, BIC, and RMSEA for Model Selection in Structural Equation Modeling. *Psychometrika*, *82*(2), 407–426. https://doi.org/10.1007/s11336-017-9572-y
Hughes, D. J. (2018). Psychometric Validity: Establishing the Accuracy and Appropriateness of Psychometric Measures. Em P. Irwing, T. Booth, & D. J. Hughes (Org.), *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development*. John Wiley & Sons Ltd.
Jackson, D. L., Gillaspy, J. A., & Purc-Stephenson, R. (2009). Reporting Practices in Confirmatory Factor Analysis: An Overview and Some Recommendations. *Psychological Methods*, *14*(1). https://doi.org/10.1037/a0014694
Jak, S., Jorgensen, T. D., Verdam, M. G. E., Oort, F. J., & Elffers, L. (2021). Analytical Power Calculations for Structural Equation Modeling: A Tutorial and Shiny App. *Behavior Research Mehods*, *53*, 1385–1406. https://doi.org/10.3758/s13428-020-01479-0/Published
Jobst, L. J., Bader, M., & Moshagen, M. (2023). A Tutorial on Assessing Statistical Power and Determining Sample Size for Structural Equation Models. *Psychological Methods*, *28*(1), 207–221. https://doi.org/10.1037/met0000423
Jones, E. A., Wind, S. A., Tsai, C.-L., & Ge, Y. (2023). Comparing Person-Fit and Traditional Indices Across Careless Response Patterns in Surveys. *Appl. Psychol. Meas.*, *47*(5-6), 365–385. https://doi.org/10.1177/01466216231194358
Jorgensen, T. D., Kite, B. A., Chen, P.-Y., & Short, S. D. (2018). Permutation randomization methods for testing measurement equivalence and detecting differential item functioning in multiple-group confirmatory factor analysis. *Psychological Methods*, *23*(4), 708–728. https://doi.org/10.1037/met0000152

Kalkbrenner, M. T. (2023). Alpha, Omega, and H Internal Consistency Reliability Estimates: Reviewing These Options and When to Use Them. *Counseling Outcome Research and Evaluation*, *14*(1), 77–88. https://doi.org/10.1080/21501378.2021.1940118

Kathawalla, U.-K., Silverstein, P., & Syed, M. (2021). Easing Into Open Science: A Guide for Graduate Students and Their Advisors. *Collabra: Psychology*, *7*(1), 18684. https://doi.org/10.1525/collabra.18684

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., IJzerman, H., Nilsonne, G., Vanpaemel, W., & Frank, M. C. (2018). A Practical Guide for Transparency in Psychological Science. *Collabra: Psychology*, *4*(1), 20. https://doi.org/10.1525/collabra.158

Kline, R. B. (2016). *Principles and Pratice of Structural Equation Modeling*. The Guilford Press.

Kline, R. B. (2023). *Principles and Pratice of Structural Equation Modeling* (Fifth Edition). The Guilford Press.

Kyriazos, T. A. (2018). Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psychology*, *09*(08), 2207–2230. https://doi.org/10.4236/psych.2018.98126

Kyriazos, T., & Poga, M. (2023). Dealing with Multicollinearity in Factor Analysis: The Problem, Detections, and Solutions. *Open Journal of Statistics*, *13*(3), 404–424. https://doi.org/10.4236/ojs.2023.133020

Lai, K. (2020). Confidence Interval for RMSEA or CFI Difference Between Nonnested Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 16–32. https://doi.org/10.1080/10705511.2019.1631704

Lai, K. (2021a). Fit Difference Between Nonnested Models Given Categorical Data: Measures and Estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(1), 99–120. https://doi.org/10.1080/10705511.2020.1763802

Lai, K. (2021b). Using Information Criteria Under Missing Data: Full Information Maximum Likelihood Versus Two-Stage Estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(2), 278–291. https://doi.org/10.1080/10705511.2020.1780925

Lei, P.-W., & Shiverdecker, L. K. (2020). Performance of Estimators for Confirmatory Factor Analysis of Ordinal Variables with Missing Data. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(4), 584–601. https://doi.org/10.1080/10705511.2019.1680292

Leite, W. L., Bandalos, D. L., & Shen, Z. (2023). Simulation Methods in Structural Equation Modeling. Em R. H. Hoyle (Org.), *Handbook of Structural Equation Modeling*. The Guilford Press.

Li, C. H. (2016a). Confirmatory Factor Analysis with Ordinal Data: Comparing Robust Maximum Likelihood and Diagonally Weighted Least Squares. *Behavior Research Methods*, *48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Li, C. H. (2016b). The Performance of ML, DWLS, and ULS Estimation with Robust Corrections in Structural Equation Models with Ordinal Variables. *Psychological Methods*, *21*(3), 369–387. https://doi.org/10.1037/met0000093

Limongi, R., & Rogers, P. (2025a). Open Science in Three Acts: Foundations, Practice, and Implementation - First Act. *BAR - Brazilian Administration Review*, *22*(1), e250079. https://doi.org/10.1590/1807-7692bar2025250079

Limongi, R., & Rogers, P. (2025b). Open Science in Three Acts: Foundations, Practice, and Implementation - Second Act. *BAR - Brazilian Administration Review*, *22*(2),

e250116. https://doi.org/10.1590/1807-7692bar2025250116
Lin, L.-C., & Yao, G. (2022). Validation of the Factor Structure of the WHOQOL-BREF Using Meta-Analysis of Exploratory Factor Analysis and Social Network Analysis. *Psychological Assessment*, *34*(7), 660–670. https://doi.org/10.1037/pas0001122
Mai, R., Niemand, T., & Kraus, S. (2021). A Tailored-Fit Model Evaluation Strategy for Better Decisions about Structural Equation Models. *Technological Forecasting and Social Change*, *173*, 121142. https://doi.org/10.1016/j.techfore.2021.121142
Marcoulides, K. M., & Yuan, K. H. (2017). New Ways to Evaluate Goodness of Fit: A Note on Using Equivalence Testing to Assess Structural Equation Models. *Structural Equation Modeling*, *24*(1), 148–153. https://doi.org/10.1080/10705511.2016.1225260
Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. (2004). Why Multicollinearity Matters: A Reexamination of Relations Between Self-Efficacy, Self-Concept, and Achievement. *Journal of Educational Psychology*, *96*(3), 518–522. https://doi.org/10.1037/0022-0663.96.3.518
Martins, H. C. (2021). Tutorial-Articles: The Importance of Data and Code Sharing. *Revista de Administração Contemporânea*, *25*(1), e200212. https://doi.org/10.1590/1982-7849rac2021200212
Maydeu-Olivares, A., Fairchild, A. J., & Hall, A. G. (2017). Goodness of Fit in Item Factor Analysis: Effect of the Number of Response Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(4), 495–505. https://doi.org/10.1080/10705511.2017.1289816
McNeish, D. (2018). Thanks Coefficient Alpha, We'll Take It from Here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144
McNeish, D. (2023a). Dynamic Fit Index Cutoffs for Factor Analysis with Likert, Ordinal, or Binary Responses. *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/tp35s
McNeish, D. (2023b). Generalizability of Dynamic Fit Index, Equivalence Testing, and Hu & Bentler Cutoffs for Evaluating Fit in Factor Analysis. *Multivariate Behavioral Research*, *58*(1), 195–219. https://doi.org/10.1080/00273171.2022.2163477
McNeish, D., & Wolf, M. G. (2023). Dynamic Fit Index Cutoffs for Confirmatory Factor Analysis Models. *Psychological Methods*, *28*(1), 61–88. https://doi.org/10.1037/met0000425
McNeish, D., & Wolf, M. G. (2024). Direct Discrepancy Dynamic Fit Index Cutoffs for Arbitrary Covariance Structure Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*(5), 835–862. https://doi.org/10.1080/10705511.2024.2308005
Meade, A. W., & Craig, S. B. (2012). Identifying Careless Responses in Survey Data. *Psychol. Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085
Mendes-Da-Silva, W. (2023). What Lectures and Research in Business Management Need to Know About Open Science. *Revista de Administração de Empresas*, *63*(4), e0000–0033. https://doi.org/10.1590/s0034-759020230408x
Moshagen, M., & Bader, M. (2024). semPower: General Power Analysis for Structural Equation Models. *Behav. Res. Methods*, *56*(4), 2901–2922. https://doi.org/10.3758/s13428-023-02254-7
Mosqueira Taipe, C. M., Amador Campos, J. A., Krieger, V., Miasso, A. I., Molina, N. P. F. M., & dos Santos, P. L. (2026). Systematic Review of the Factor Structure of WHOQOL-bref in Adult Population. *Quality of Life Research*, *35*(17), 1–18. https://doi.org/10.1007/s11136-025-04115-6
Nalbantoğlu-Yılmaz, F. (2019). Comparison of Different Estimation Methods Used in Confirmatory Factor Analyses in Non-Normal Data: A Monte Carlo Study. *International Online Journal of Educational Sciences*, *11*(4). https://doi.org/10.15345/iojes.

2019.04.010
Niemand, T., & Mai, R. (2018). Flexible Cutoff Values for Fit Indices in the Evaluation of Structural Equation Models. *Journal of the Academy of Marketing Science*, *46*(6), 1148–1172. https://doi.org/10.1007/s11747-018-0602-9
Norman, G. (2010). Likert Scales, Levels of Measurement and the "Laws" of Statistics. *Adv. Health Sci. Educ.*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y
Nye, C. D. (2022). Reviewer Resources: Confirmatory Factor Analysis. *Organizational Research Methods*, 109442812211205. https://doi.org/10.1177/10944281221120541
Pilcher, N., & Cortazzi, M. (2023). 'Qualitative' and 'quantitative' Methods and Approaches across Subject Fields: Implications for Research Values, Assumptions, and Practices. *Quality & Quantity*. https://doi.org/10.1007/s11135-023-01734-4
Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of Method Bias in Social Science Research and Recommendations on How to Control It. *Annual Review of Psychology*, *63*(1), 539–569. https://doi.org/10.1146/annurev-psych-120710-100452
Pornprasertmanit, S., Miller, P., Jorgensen, T. D., & Corbin, Q. (2022). *Simsem: SIMulated Structural Equation Modeling*. R package. www.simsem.org
Preacher, K. J., & Yaremych, H. E. (2023). Model Selection in Structural Equation Modeling. Em R. H. Hoyle (Org.), *Handbook of Structural Equation Modeling*. The Guilford Press.
Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004
Reeves, T. D., & Marbach-Ad, G. (2016). Contemporary Test Validity in Theory and Practice: A Primer for Discipline-Based Education Researchers. *CBE Life Sciences Education*, *15*(1). https://doi.org/10.1187/cbe.15-08-0183
Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When Can Categorical Variables Be Treated as Continuous? A Comparison of Robust Continuous and Categorical SEM Estimation Methods under Suboptimal Conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315
Rios, J., & Wells, C. (2014). Validity Evidence Based on Internal Structure. *Psicothema*, *26*(1), 108–116. https://doi.org/10.7334/psicothema2013.260
Robitzsch, A. (2020). Why Ordinal Variables Can (Almost) Always Be Treated as Continuous Variables: Clarifying Assumptions of Robust Continuous and Ordinal Factor Analysis Estimation Methods. *Frontiers in Education*, *5*. https://doi.org/10.3389/feduc.2020.589965
Robitzsch, A. (2022). On the Bias in Confirmatory Factor Analysis When Treating Discrete Variables as Ordinal Instead of Continuous. *Axioms*, *11*(4). https://doi.org/10.3390/axioms11040162
Rogers, P. (2022). Best Practices for Your Exploratory Factor Analysis: A Factor Tutorial. *Revista de Administração Contemporânea*, *26*(6). https://doi.org/10.1590/1982-7849rac2022210085.en
Rogers, P. (2024). Best Practices for Your Confirmatory Factor Analysis: A JASP and Lavaan Tutorial. *Behavior Research Methods*. https://doi.org/10.3758/s13428-024-02375-7
Rogers, P., & Limongi, R. (2025). Open Science in Three Acts: Foundations, Practice, and Implementation - Third Act. *BAR - Brazilian Administration Review*, *22*(3), e250162. https://doi.org/10.1590/1807-7692bar2025250162
Rosseel, Y. (2012). Lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rosseel, Y. (2023). *The Lavaan Tutorial*. Department of Data Analysis, Ghent University.
Schumacker, R. E., Wind, S. A., & Holmes, L. F. (2021). Resources for Identifying Measurement Instruments for Social Science Research. *Measurement: Interdisciplinary Research and Perspectives*, *19*(4), 250–257. https://doi.org/10.1080/15366367.2021.1950486
Shek, D. T. L., & Yu, L. (2014). Use of Structural Equation Modeling in Human Development Research. *International Journal on Disability and Human Development*, *13*(2), 157–167. https://doi.org/10.1515/ijdhd-2014-0302
Shi, D., & Maydeu-Olivares, A. (2020). The Effect of Estimation Methods on SEM Fit Indices. *Educational and Psychological Measurement*, *80*(3), 421–445. https://doi.org/10.1177/0013164419885164
Shi, D., Maydeu-Olivares, A., & Rosseel, Y. (2020). Assessing Fit in Ordinal Factor Analysis Models: SRMR vs. RMSEA. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 1–15. https://doi.org/10.1080/10705511.2019.1611434
Sireci, S. G., & Sukin, T. (2013). Test Validity. Em *APA Handbook of Testing and Assessment in Psychology, Vol. 1: Test Theory and Testing and Assessment in Industrial and Organizational Psychology.* (p. 61–84). American Psychological Association. https://doi.org/10.1037/14047-004
Skevington, S. M., & Epton, T. (2018). How Will the Sustainable Development Goals Deliver Changes in Well-Being? A Systematic Review and Meta-Analysis to Investigate Whether WHOQOL-BREF Scores Respond to Change. *BMJ Glob. Health*, *3*. https://doi.org/10.1136/bmjgh-2017-000609
Steedle, J. (2018). *Detecting Inattentive Responding on a Psychosocial Measure of College Readiness*. ACT.org.
Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements. *Frontiers in Psychology*, *7*(MAY). https://doi.org/10.3389/fpsyg.2016.00769
Ulitzsch, E., Pohl, S., Khorramdel, L., Kroehne, U., & Davier, M. von. (2022). A Response-Time-Based Latent Response Mixture Model for Identifying and Modeling Careless and Insufficient Effort Responding in Survey Data. *Psychometrika*, *87*(2), 593–619. https://doi.org/10.1007/s11336-021-09817-7
Urbina, S. (2014). *Essentials of Psychological Testing*. John Wiley & Sons.
Wang, Y. A. (2023). *How to Conduct Power Analysis for Structural Equation Models: A Practical Primer* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/4n3uk
Wang, Y. A., & Rhemtulla, M. (2021). Power Analysis for Parameter Estimation in Structural Equation Modeling: A Discussion and Tutorial. *Advances in Methods and Practices in Psychological Science*, *4*(1), 1–17. https://doi.org/10.1177/2515245920918253
Ward, M. K., & Meade, A. W. (2023). Dealing with Careless Responding in Survey Data: Prevention, Identification, and Recommended Best Practices. *Annu. Rev. Psychol.*, *74*(Volume 74, 2023), 577–596. https://doi.org/10.1146/annurev-psych-040422-045007
West, S. G., Wu, W., McNeish, D., & Savord, A. (2023). Model Fit in Structural Equation Modeling. Em R. H. Hoyle (Org.), *Handbook of Structural Equation Modeling*. The Guilford Press.
Westland, C. J. (2010). Lower Bounds on Sample Size in Structural Equation Modeling. *Electronic Commerce Research and Applications*, *9*(6). https://doi.org/10.1016/j.elerap.2010.07.003
Whittaker, T. A., & Schumacker, R. E. (2022). *A Beginner's Guide to Structural Equation Modeling* (Fifth Edition). Routledge Taylor & Francis Group.

Widodo, E. (2018). Some Notes on the Contemporary Views of Validity in Psychological and Educational Assessment. *Advances in Social Science, Education and Humanities Research*, *231*, 732–734. https://doi.org/10.3968/8877

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in Structural Equation Modeling with Ordered Categorical Data: The Story They Tell Depends on the Estimation Methods. *Behavior Research Methods*, *51*(1), 409–428. https://doi.org/10.3758/s13428-018-1055-2

Yang, Y., & Liang, X. (2013). Confirmatory Factor Analysis under Violations of Distributional and Structural Assumptions. *Int. J. Quantitative Research in Education*, *1*(1), 61–84.

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory Factor Analysis of Ordinal Variables with Misspecified Models. *Structural Equation Modeling*, *17*(3), 392–423. https://doi.org/10.1080/10705511.2010.489003

Yoshitake, N., Sun, Y., Sugawara, M., Matsumoto, S., Sakai, A., Takaoka, J., & Goto, N. (2015). The Psychometric Properties of the WHOQOL-BREF in Japanese Couples. *Health Psychology Open*, *2*(2), 1–9. https://doi.org/10.1177/2055102915598089

Yuan, K. H., Chan, W., Marcoulides, G. A., & Bentler, P. M. (2016). Assessing Structural Equation Models by Equivalence Testing With Adjusted Fit Indexes. *Structural Equation Modeling*, *23*(3), 319–330. https://doi.org/10.1080/10705511.2015.1065414

Zhang, Z., & Yuan, K.-H. (2018). *Practical Statistical Power Analysis Using Webpower and R*. ISDSA Press. https://doi.org/10.35566/power