



---

## AVALIAÇÃO DE INSTRUMENTOS PSICOMÉTRICOS NO SOFTWARE R: ANÁLISES DE CONFIABILIDADE E VALIDADE

THAIS ACCIOLY BACCARO  
RODRIGO LIBANEZ MELAN  
SAULO FABIANO AMÂNCIO VIEIRA

Artigo convidado e aceito para publicação em 29/08/2022  
DOI: 10.5433/2318-9223.2021v9n1p92-110

### RESUMO

A validação de instrumentos psicométricos enquanto ferramentas adequadas para se mensurar fenômenos atinentes aos indivíduos requer emprego de métodos e técnicas para se apreciar a sua consistência, ou seja, confiabilidade, bem como para se analisar a proximidade com a medida à qual se relaciona, também denominada validade. Apesar do desenvolvimento de maneiras de fazê-lo, impulsionadas pela evolução tecnológica, elas ainda se encontram fragmentadas e, por vezes, de difícil e/ou custosa implementação. Este artigo metodológico trata da realização de procedimentos para validação de instrumentos psicométricos, indicando seus comandos e exemplificando os resultados no software estatístico R. Portanto, visa à consolidação, simplificação e redução dos esforços necessários para tanto, considerando, ainda, que o software é livre para uso.

**PALAVRAS-CHAVE.** Psicometria. Análise Fatorial Confirmatória. Avaliação Psicológica.

### EVALUATION OF PSYCHOMETRIC INSTRUMENTS IN SOFTWARE R: RELIABILITY AND VALIDITY ANALYSIS

### ABSTRACT

Validation of psychometric instruments as adequate tools to measure phenomena related to individuals requires the use of methods and techniques to assess their consistency, that is, reliability, and to analyze the proximity to the measure to which it relates, also called validity. Despite developing ways to do this, driven by technological developments, they are still fragmented and, at times, complex and costly to implement. This methodological paper deals with the realization of procedures for validation of psychometric instruments,

indicating their commands and exemplifying results in the statistical software R. Therefore, it aims at consolidating, simplifying, and reducing efforts necessary for this, considering still that the software is free for use.

**KEYWORDS.** Psychometry. Confirmatory Factor Analysis. Psychological Assessment.

## INTRODUÇÃO

Os instrumentos psicométricos são importantes ferramentas em áreas que envolvem conhecimento acerca dos indivíduos e requerem esforços para adequação de seus atributos métricos, ou seja, para o processo de validação. Existe ampla literatura a respeito da avaliação de tais atributos, embora nem sempre trate consensualmente sobre os métodos e ferramentas adequadas a tanto. Durante o século passado, as discussões sobre o tema (Pacico & Hutz, 2015) expressaram desafios à utilização das técnicas de análise de instrumentos, sem contar os obstáculos postos pela complexidade estatística envolvida na validação.

O artigo, além de oferecer tratamento objetivo à análise de instrumentos psicométricos, disponibiliza exemplos claros de sua aplicação no software estatístico R. Este consiste em uma poderosa ferramenta, além de possuir uso livre, sendo encontrado em diretório na internet (R Core Team, 2020). Alguns pacotes não nativos são necessários ao desenvolvimento das análises aqui demonstradas, quais sejam “MBESS” (Kelley, 2020), “irrCAC” (Gwet, 2019b), “MVN” (Korkmaz et al., 2014) e “lavaan” (Rosseel, 2012). Eles podem ser instalados por meio do comando “install.packages(“nome\_do\_pacote”, dep=T)”.

A validação é abordada no artigo por meio dos conceitos necessários ao seu desempenho, ou seja, a confiabilidade e a validade. O primeiro refere-se à proximidade entre medidas em situações iguais ou semelhantes o segundo, à proximidade com o valor real do item avaliado (Babbie, 2016; Pasquali, 2009).

## ANÁLISE DE CONFIABILIDADE

Se um instrumento psicométrico resulta em medidas iguais ou próximas quando aplicado a um mesmo objeto, diz-se que ele é confiável ou fidedigno (Babbie, 2016; Zanon & Hauck Filho, 2015), ou ainda, que possui consistência (Fitzner, 2007). Mais especificamente, a confiabilidade ocorre quando o instrumento aplicado aos mesmos sujeitos em distintas situações ou instrumentos diferentes e que se equivalem utilizados no mesmo sujeito e mesma situação apresentam resultados semelhantes (Pasquali, 2009). Uma das medidas bastante disseminadas para tanto é o Alfa de Cronbach ( $\alpha$ ) (Cronbach & Shavelson, 2004), compreendido enquanto o grau de coerência interna de um questionário.

A utilização de  $\alpha$  requer tau-equivalência do modelo em que o instrumento se baseia e ausência de correlação entre os erros dos itens que o compõem (Teo & Fan, 2013), o que pode torná-lo problemático como uma medida de confiabilidade. Outrossim, elevados valores do coeficiente podem indicar inadequações e/ou falsas percepções a respeito do instrumento, o que provém de sua dependência do número de itens e de escalas, bem como da redundância interna que números muito altos de  $\alpha$  podem sugerir (Streiner et al., 2015). Assim, quando o modelo não atende ou não se conhece a tau-equivalência e/ou a correlação entre os erros de seus itens, preza-se pelo cômputo e análise do coeficiente ômega ( $\omega$ ). Existem, ainda, outras abordagens ao  $\alpha$ , como a variância média extraída e o índice de confiabilidade composta (Valentini & Damásio, 2016).

No R (R Core Team, 2020), o pacote “MBESS” permite o cálculo de  $\alpha$  e  $\omega$  (Kelley, 2020), e pode ser instalado pelo comando “install.packages(“MBESS”, dep=T)”. Para se exemplificar os cálculos e resultados, utilizou-se o questionário de mensuração do pensamento reflexivo (QMPR), desenvolvido por Kember et al. (2000) e adaptado e validado por Melan (2018). O QMPR abarca, originalmente, quatro constructos: *habitual action* (HA), *understanding* (U), *reflection* (R) e *critical reflection* (CR), os quais são mensurados, cada um, por quatro indicadores (afirmações) e estes, por sua vez, apresentam quatro possibilidades de respostas em escala Likert e uma a ser utilizada quando a concordância ou discordância não é aplicável (Kember et al., 2000). Os comandos e resultados constam na **Figura 1**.

Os dados de entrada no software R (R Core Team, 2020) estão na planilha eletrônica “matriz\_original\_dados\_QMPR.csv”, sendo que cada coluna apresenta um indicador (afirmação) e as linhas, respostas ao questionário. Destaca-se que a planilha contém cabeçalho e está em formato “csv”, o que justifica a especificação “;” para o separador, “,” para os decimais e “header=T”. Além do valor geral de  $\alpha$  e  $\omega$ , observa-se que os coeficientes se encontram calculados em todas escalas do QMPR, como por exemplo, em relação àquela de HA por meio dos indicadores 1, 5, 9 e 13 (vide área em destaque na **Figura 1**). Esses números referem-se, respectivamente, às colunas em que estão A1, A5, A9 e A13 na planilha de entrada (para a estrutura completa do QMPR, vide **Figura 5**). Os valores de HA e R, bem como do QMPR (geral), mostraram-se abaixo de 0,7, sugerindo confiabilidade inadequada (Hair, Babin et al., 2005).

No próximo tópico, serão abordados conceitos de validade e maneiras de analisá-la, bem como a aplicação destas no software R (R Core Team, 2020).

FIGURA 1

COMANDOS E RESULTADOS PARA CÁLCULO DE  $\alpha$  E  $\omega$ 

```
> dados <- read.table("matriz_original_dados_QMPR.csv", sep=";", dec=".", header=T)
> library("MBESS")
> ci.reliability(data=dados[,c(1,5,9,13)], type="alpha", interval.type="none")$est
[1] 0.489718
> ci.reliability(data=dados[,c(2,6,10,14)], type="alpha", interval.type="none")$est
[1] 0.7687705
> ci.reliability(data=dados[,c(3,7,11,15)], type="alpha", interval.type="none")$est
[1] 0.6433303
> ci.reliability(data=dados[,c(4,8,12,16)], type="alpha", interval.type="none")$est
[1] 0.7288189
>
> ci.reliability(data=dados, type="alpha", interval.type="none")$est
[1] 0.6256323
>
> ci.reliability(data=dados[,c(1,5,9,13)], type="omega", interval.type="none")$est
[1] 0.4965834
> ci.reliability(data=dados[,c(2,6,10,14)], type="omega", interval.type="none")$est
[1] 0.7756282
> ci.reliability(data=dados[,c(3,7,11,15)], type="omega", interval.type="none")$est
[1] 0.6545643
> ci.reliability(data=dados[,c(4,8,12,16)], type="omega", interval.type="none")$est
[1] 0.7301967
>
> ci.reliability(data=dados, type="omega", interval.type="none")$est
[1] 0.4976249
```

FONTE: ADAPTADO DE KELLEY (2020) E R CORE TEAM (2020).

## ANÁLISE DE VALIDADE

Em um sentido moderno, a validade pode ser entendida como congruência entre o valor obtido pela mensuração e a característica mensurada do objeto (Pasquali, 2009). Não há identificação perfeita entre ambos, ou seja, o instrumento psicométrico captará, no máximo, algo próximo da característica real, embora a partir de certo nível em que isso ocorra, as medidas poderão ser úteis a quem delas se valer (Babbie, 2016). A validade se expressa em quatro tipos, quais sejam de face, de critério, de constructo e de conteúdo (Babbie, 2016; Souza et al., 2017), sendo que eles devem ser alcançados para que o instrumento possua capacidade de mensuração satisfatória.

A validade de face é atingida quando um indicador (medida) aparentemente, isto é, em uma primeira análise, mais imediata, possui relação com o constructo a que se propõe medir (Babbie, 2016). Sua avaliação se dá subjetivamente (Fitzner, 2007), o que desvela a relevância dos juízes (experts) nessa função (Hardesty & Bearden, 2004). Esse tipo de validade pode ser apreciado classificando as opiniões deles quanto à adequação do indicador nas categorias desenvolvidas por Zaichkowsky, “claramente representativo”, “moderadamente representativo” e “não representativo” (Hardesty & Bearden, 2004;

Zaichkowsky, 1985). Na prática, uma das maneiras de se proceder a tal avaliação é calcular, por meio de índices, o nível de concordância dos juízes sobre a pertinência do indicador, solicitando que preencham o instrumento. Entre os índices mais utilizados, está o coeficiente Kappa, embora ele possa apresentar baixos valores no contexto de alta concordância observada, o que se deve à forma de distribuição das respostas nas categorias utilizadas para classificação (Delgado & Tibau, 2019; Gwet, 2002). Uma alternativa ao Kappa é o coeficiente de classificação (*agreement coefficient*), doravante AC, desenvolvido por Kilem L. Gwet, o qual é calibrado para não incorrer no referido problema (Gwet, 2002, 2008).

Para se exemplificar o cálculo de AC, suponha que seis juízes avaliaram a adequação dos indicadores de 1 a 12 (por exemplo, afirmações relacionadas a um constructo). A eles foram dadas cinco opções, devendo indicar o número 5, 4, 3, 2 e 1, nesta ordem, se o indicador é perfeitamente, muito, moderadamente, pouco ou não representativo do constructo. Os resultados estão dispostos na **Tabela 1** (cada linha refere-se a um indicador).

**TABELA 2**  
RESULTADOS DA CLASSIFICAÇÃO DE QUATRO JUÍZES PARA 12 INDICADORES

Juiz 1	Juiz 2	Juiz 3	Juiz 4
1	1		1
2	2	3	2
3	3	3	3
3	3	3	3
2	2	2	2
1	2	3	4
4	4	4	4
1	1	2	1
2	2	2	2
	5	5	5
		1	1
		3	

FONTE: ADAPTADO DE GWET (2019A).

Os espaços em branco representam classificações não realizadas, o que expressa a flexibilidade das entradas para o cálculo de AC (Gwet, 2019a). Deve-se salvar os dados da **Tabela 1** em uma planilha eletrônica em formato “csv” (aqui ela é denominada “resultados\_juizes\_para\_ac”), embora outros formatos também possam ser utilizados. O pacote “irrCAC” (Gwet, 2019b) é requerido ao cálculo de AC e, assim, deve ser instalado por meio do comando “install.packages (“irrCAC”, dep=T)”. Na **Figura 2**, os cálculos e resultados são apresentados e indicam AC de 0,77544, valor que, se fosse tomado para o coeficiente Kappa, representaria concordância substancial entre os juízes (Landis & Koch, 1977). Ressalta-se, ainda, que ele é significativo em níveis de confiança maiores do que 99%, considerando p-valor de 0,000208721.

FIGURA 2  
COMANDOS E RESULTADOS ASSOCIADOS À AC

```
> dados <- read.table("resultados_juizes_para_ac.csv", sep=";", dec=".", header=T)
> library(irrCAC)
> acl <- gwet.acl.raw(dados)$est
> acl
  coeff.name      pa      pe coeff.val coeff.se conf.int      p.value      w.name
1      AC1 0.8181818 0.1903212  0.77544  0.14295 (0.461,1) 0.000208721 unweighted
```

FONTE: ADAPTADO DE GWET (2019A, 2019B) E R CORE TEAM (2020).

Após a apreciação da validade de face, procede-se à de critério, a qual diz respeito à capacidade de o score no instrumento em avaliação deduzir posições do respondente em uma medida externa (critério), ou seja, seu desempenho específico nesta (Cohen et al., 2014; Pasquali, 2009). Esse tipo de validade pode ocorrer em duas formas, a preditiva, quando as medidas no instrumento e no critério são tomadas em momentos distintos, e a concorrente, se as obter aproximadamente ao mesmo tempo (Pacico & Hutz, 2015). Todavia, não é importante, no que concerne ao aspecto técnico, a forma de obtenção dos dados, mas sim a adequação do critério (Pasquali, 2009). Assim, ele deve ser relevante, válido e não contaminado: respectivamente apresentar relação com o instrumento, ter validade assegurada e não possuir indicadores contidos no instrumento em avaliação (Pacico & Hutz, 2015). Um coeficiente de correlação fornece base para analisar a validade de critério, especificamente por meio do teste da hipótese de associação entre medida e critério (Pacico & Hutz, 2015; Souza et al., 2017).

A ilustração do cálculo de correlação no software R (R Core Team, 2020) demandará a consideração dos dados da **Tabela 2**. Nele é possível observar duas colunas, a das pontuações no instrumento em análise (Instrumento) e no de referência (Critério), cujos pontos servirão à classificação no critério (cada linha representa um respondente). O coeficiente utilizado é o de Spearman, pois este se adere a contextos em que as variáveis são quantitativas contínuas e/ou discretas, bem como naqueles em que a normalidade distributiva dos dados não se verifica para uma das variáveis ao menos (Lehman et al., 2013). A **Tabela 2** foi salva em planilha eletrônica sob o nome “resultados\_instrumentos\_para\_correlacao.csv”.

Por meio da **Figura 3** observa-se que o resultado do coeficiente de Spearman foi 0,8952828, enquanto seu nível de significância apresentou-se maior do que 99% face ao p-valor de 0,00003538. Não há consenso sobre números satisfatórios de correlação (Pacico & Hutz, 2015), embora quanto mais próximo de + 1 ou -1, nesta ordem, das associações positiva e negativa perfeitas, maior a correlação.

**TABELA 2**  
PONTUAÇÕES NO INSTRUMENTO EM ANÁLISE E NO DE REFERÊNCIA (CRITÉRIO)

Instrumento	Crítério
10	2
15	4
12	2
20	8
10	3
11	3
18	9
13	3
19	8
8	1
14	5
20	10
11	2

FONTE: ELABORAÇÃO PRÓPRIA.

A validade de constructo prossegue a de critério e caracteriza-se na correspondência entre medida (constructo) e variáveis associadas (indicadores) assim como previsto no arcabouço teórico que embasou o instrumento (Babbie, 2016). Ela pode ser aferida pela técnica de grupos conhecidos e por subtipos de validade, como a convergente, discriminante, estrutural e transcultural (Souza et al., 2017). Neste trabalho o foco recairá sobre a avaliação de estrutura por meio da análise fatorial confirmatória (AFC), a qual é uma vertente da modelagem de equações estruturais (MEE). Ressalta-se que, com o avanço da tecnologia da informação e da capacidade de processamento dos computadores, as técnicas estatísticas multivariadas se popularizaram, levando à disseminação da MEE na ciência psicológica (Pilati & Laros, 2007). A aplicabilidade da AFC no software livre R (R Core Team, 2020) corrobora a relevância da técnica.

**FIGURA 3**  
COMANDOS E RESULTADOS DA CORRELAÇÃO DE SPEARMAN

```
> dados <- read.table("resultados_instrumentos_para_correlacao.csv", sep=";", dec=".", header=T)
> cor.test(Instrumento, Critério, method="spearman")

Spearman's rank correlation rho

data: Instrumento and Critério
S = 38.117, p-value = 3.538e-05
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.8952828

Warning message:
In cor.test.default(Instrumento, Critério, method = "spearman") :
Impossível calcular o valor exato de p com empates
```

FONTE: ADAPTADO DE R CORE TEAM (2020).



A MEE, e, portanto, a AFC, permite que se avalie relações de dependência (típicas de regressões), proceda-se à análise fatorial (mensurando conceitos não medidos diretamente, ou seja, constructos) e as realize concomitante e multiplamente (Hair, Anderson et al., 2005). Para se exemplificar a implementação da técnica, retomou-se o questionário de mensuração do pensamento reflexivo (QMPR), desenvolvido por (Kember et al., 2000) e adaptado e validado por Melan (2018). Antes de se empreender a técnica, alguns pressupostos da distribuição de dados e da relação entre eles se fazem necessários, conforme indicado a seguir.

O primeiro deles diz respeito à matriz de entrada de dados, ou seja, a planilha com as respostas às 16 afirmações do QMPR. Ela deve ser positivamente definida, o que implica não-singularidade (possui inversa e, portanto, seu determinante é diferente de 0), autovalores positivos e módulos das covariâncias entre seus indicadores (intensidade de concordância e discordância para cada afirmação) abaixo ou igual ao número que resulta da expressão  $\sqrt{S_X^2 * S_Y^2}$  (Kline, 2016; Santos, 2017). Os termos “S<sup>2</sup>X” e “S<sup>2</sup>Y” referem-se à variância de dois indicadores quaisquer X e Y. O cálculo deverá ser realizado e avaliado para todos pares deles. Além da definição positiva para a matriz de entrada de dados, deve-se verificar se os indicadores possuem altas colinearidades; se existem outliers; e a normalidade multivariada quanto à distribuição dos dados (Kline, 2016). Destaca-se, ainda, a necessidade de se escolher o tipo de matriz, sendo que no caso da MEE, por se tratar de técnica para avaliação de relações causais, recomenda-se a de variância-covariância (Hair, Babin et al., 2005). Portanto, considerando-se que o arquivo “matriz\_original\_dados\_QMPR.csv” contém as respostas ao QMPR, em que cada coluna apresenta valores referentes a uma afirmação, pontuadas de 1 (“discordo totalmente”) a 5 (“concordo totalmente”), a **Figura 3** traz os comandos no R (R Core Team, 2020) e seus resultados.

O determinante da matriz de entrada de dados, cujo valor foi de 0,150234 (em destaque na **Figura 4**), indica que ela possui inversa. Os autovalores são todos positivos, conforme indicado na área que resultou do comando “eigen(cov(dados), only.values=T)”. Em se tratando da diferença entre o produto da raiz das variâncias de cada indicador ( $\sqrt{S_X^2 * S_Y^2}$ ) e o módulo de suas covariâncias, abarcada pelo objeto “matriz\_diferenca”, ela é exposta (parcialmente, devido à conveniência de leiaute) na última área em destaque. Os valores são positivos e, em conjunto dos outros resultados, expressam a definição positiva da matriz de entrada de dados.

As colinearidades podem ser analisadas por meio do fator de inflação da variância (FIV), sendo altas quando este superar o valor de 10 (Kline, 2016). Sua fórmula consiste na razão entre 1 e  $(1 - R^2)$ , com  $R^2$  representando o coeficiente de determinação para cada par de indicadores (Kline, 2016). Já a observação dos *outliers* se dá pelo valor da distância de Mahalanobis ao quadrado dividido pelos graus de liberdade (GL) do modelo do instrumento,



constatando a sua presença quando essa razão for maior do que 3 (Hair et al., 2010). GL é calculado como segue (Hair et al., 2010; Kline, 2016):

$$GL = \frac{1}{2} [(p)(p + 1)] - t \quad (1)$$

FIGURA 4

COMANDOS E RESULTADOS PARA AVALIAÇÃO DE DEFINIÇÃO POSITIVA DA MATRIZ DE ENTRADA DE DADOS

```
> dados <- read.table("matriz_original_dados_QMPR.csv", sep=";", dec=".", header=T)
> det(cov(dados))
[1] 0.150234
> eigen(cov(dados), only.values=T)
$values
[1] 3.8857808 2.6801572 1.5939960 1.3436438 1.2709175 1.0693241 0.9583970
[8] 0.7811434 0.7005462 0.6320331 0.6091943 0.5304053 0.5227408 0.4832213
[15] 0.4692091 0.3904056

$vectors
NULL

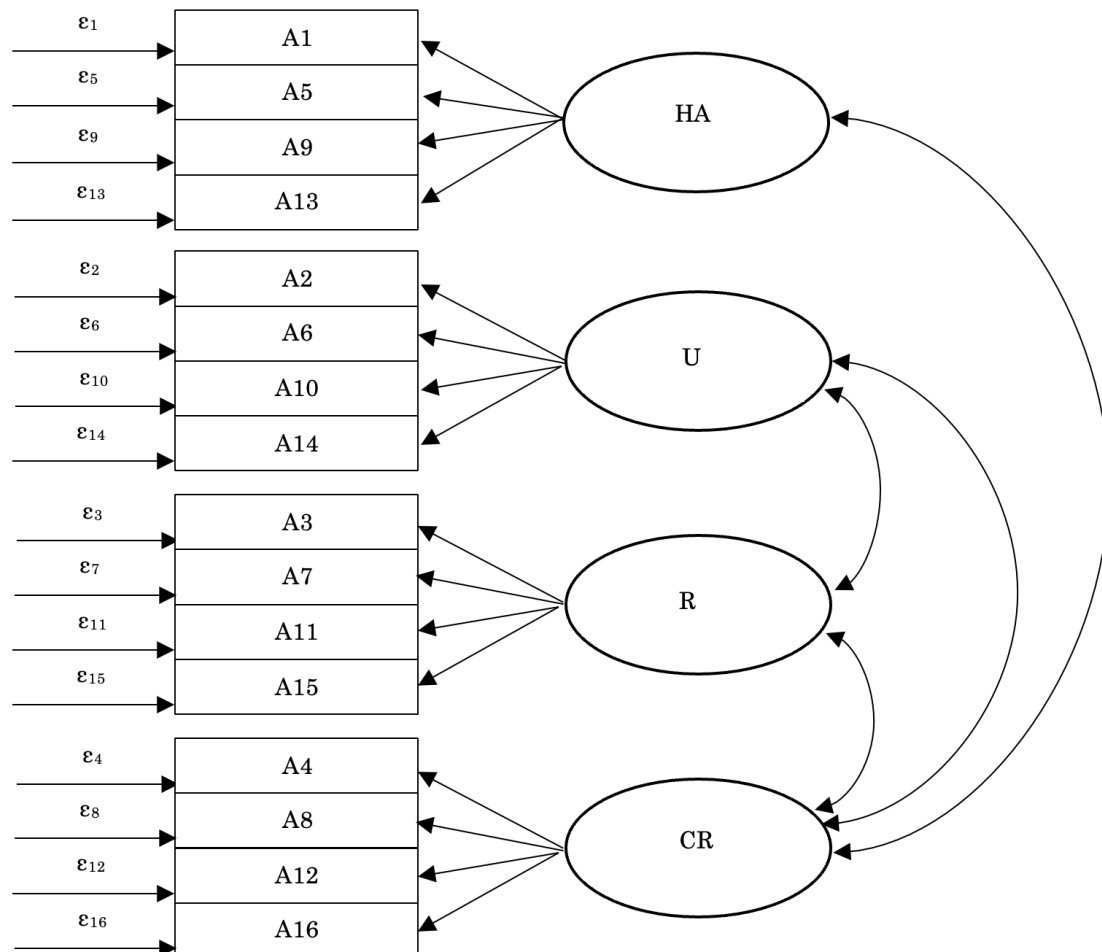
> matriz_1 <- matrix(diag(var(dados)), nrow=ncol(dados), ncol=ncol(dados), byrow=T)
> matriz_2 <- matrix(diag(var(dados)), nrow=ncol(dados), ncol=ncol(dados))
> matriz_diferenca <- (matriz_1*matriz_2)^(1/2) - abs(cov(dados))
> matriz_diferenca
```

	A1	A2	A3	A4	A5	A6	A7
A1	0.0000000	0.8839940	0.9478820	1.2474803	1.0818248	0.9590835	0.9373638
A2	0.8839940	0.0000000	0.8519928	1.0642254	1.0277042	0.4948172	0.8954091

FONTE: ADAPTADO DE R CORE TEAM (2020).

Na Equação 1, os termos “p” e “t” referem-se, respectivamente, ao número de indicadores (afirmações) e à quantidade de parâmetros livres. Estes correspondem às estimativas da AFC: cargas (impacto expresso por um coeficiente atrelado à variável independente, ou seja, constructo, na dependente, qual seja, o indicador), erros (diferenças entre a medição e seu valor real), variâncias (dos constructos) e covariâncias (entre os constructos). Para elucidar a determinação de GL, considera-se o modelo previsto do QMPR na **Figura 5**, o qual proveio do original. São 12 cargas, sendo que, para cada constructo, HA, U, R e CR, existem quatro delas, representadas pelas setas que partem deles e encontram os indicadores. Estabeleceu-se, todavia, o valor de 1 àquelas associadas às primeiras afirmações, ação necessária para se determinar o valor dos constructos. Ademais, existem 16 erros (um para cada indicador); quatro variâncias e quatro covariâncias (uma para cada constructo), totalizando o valor de 36 para “t”. Portanto, GL será:  $GL = \frac{1}{2} [(16)(16 + 1)] - 36 = 100$

FIGURA 5  
MODELO PREVISTO DO QMPR



FONTE: ADAPTADO DE KEMBER ET AL. (2000).

Sobre a normalidade multivariada, ela pode ser objeto de análise por meio da normalidade univariada, embora de maneira aproximada. Realiza-se tal ação pelo empenho necessário à avaliação daquela e devido a ser constatados por testes de variáveis únicas muitos casos em que não há normalidade multivariada (Kline, 2016). Existem, contudo, testes de normalidade multivariada cujo uso é recomendado, como o de Henze-Zirkler e de Mardia para assimetria e para curtose (Mecklin & Mundfrom, 2005), e bastante simples de efetuar por meio do R (R Core Team, 2020). Para utilizá-los, deve-se baixar a pacote “MVN” (Korkmaz et al., 2014) e instalá-lo pelo comando “install.packages(“MVN, dep=T”)”. Assim, tendo obtido os graus de liberdade do modelo, conforme indicado anteriormente, procede-se aos cálculos associados às colinearidades, ao estudo de outliers e à normalidade.

FIGURA 6

COMANDOS E RESULTADOS PARA AVALIAÇÃO DE COLINEARIDADE ENTRE INDICADORES,  
DA EXISTÊNCIA DE OUTLIERS E DA NORMALIDADE MULTIVARIADA

```
> dados <- read.table("matriz_original_dados_QMPR.csv", sep=";", dec=".", header=T)
> FIV <- (1 - cor(dados)^2)^(-1)
> FIV
      A1      A2      A3      A4      A5      A6      A7      A8
A1      Inf 1.031057 1.004289 1.000000 1.024084 1.037216 1.015165 1.004361
A2 1.031057      Inf 1.003304 1.001864 1.009610 1.392209 1.003516 1.002717
> media_coluna <- colMeans(dados)
> matriz_covariancia <- cov(dados)
> matriz_outliers <- mahalanobis(dados, media_coluna, matriz_covariancia)/100
> matriz_outliers
[1] 0.53216569 0.13121644 0.21397124 0.11545343 0.22258467 0.17877537
[7] 0.20978228 0.12976536 0.23386047 0.33226671 0.12298441 0.09034535
> library("MVN")
> teste_muultivariado_mardia <- mvn(data = dados, mvnTest = "mardia")
> teste_muultivariado_mardia
$multivariateNormality
      Test      Statistic      p value Result
1 Mardia Skewness 1230.1400685606 2.43495534618216e-19 NO
2 Mardia Kurtosis 8.89920131917578 0 NO
3      MVN      <NA>      <NA>      NO
> teste_muultivariado_HZ <- mvn(data = dados, mvnTest = "hz")
> teste_muultivariado_HZ
$multivariateNormality
      Test      HZ      p value MVN
1 Henze-Zirkler 1.013019 0 NO
```

FONTE: ADAPTADO DE KORKMAZ ET AL. (2014) E R CORE TEAM (2020).

A primeira área em destaque na **Figura 6** corresponde aos valores de FIV, exibidos parcialmente para melhor ajuste de layout. A diagonal principal, em que constam os termos “Inf”, não é considerada na análise, pois abarca valores para variáveis individuais, ou seja, o coeficiente de determinação é 1 (implica uma razão do tipo “a/0”, com a = 1). Verificou-se que todos cálculos de FIV avaliados resultaram em números menores do que 10 e, dessa forma, as colinearidades não se mostraram altas (Kline, 2016). A próxima área em destaque (também apresentada parcialmente) indicou que a distância de Mahalanobis ao quadrado dividida pelos graus de liberdade do modelo não superou o valor 3 e, por conseguinte, os dados não contêm *outliers* (Hair et al., 2010). Por fim, os testes de Mardia e de Henze-Zirkler apontaram, todos, a ausência de normalidade multivariada, o que não inviabiliza a realização da AFC, mas implica o método de estimação utilizado (discutido adiante). Ressalva-se que a realização dos testes de Mardia e Henze-Zirkler requer a ausência de outliers (Korkmaz et al., 2014), verificada pela razão entre a distância de Mahalanobis ao quadrado e GL.

FIGURA 7  
IMPLEMENTAÇÃO DA AFC NO R

```
> dados <- read.table("matriz_original_dados_QMFR.csv", sep=";", dec=".", header=T)
> library(lavaan)
> modelo_estrutural<-'
+ HA =~ A1 + A5 + A9 + A13
+ U =~ A2 + A6 + A10 + A14
+ R =~ A3 + A7 + A11 + A15
+ CR =~ A4 + A8 + A12 + A16
+
+ HA =~ 0*U
+ HA =~ 0*R
+
> fit.modelo_estrutural <- cfa(modelo_estrutural, data=dados, estimator="MLM", se="robust")
> summary(fit.modelo_estrutural, fit.measures=T, standardized=T)
lavaan 0.6-6 ended normally after 48 iterations

Estimator                               ML
Optimization method                     NLMINB
Number of free parameters                 36

Number of observations                    251

Model Test User Model:

Test Statistic                           Standard      Robust
Degrees of freedom                        100           100
P-value (Chi-square)                      0.000         0.000
Scaling correction factor                  1.137
Satorra-Bentler correction

Model Test Baseline Model:

Test statistic                            887.684       790.526
Degrees of freedom                        120           120
P-value                                   0.000         0.000
Scaling correction factor                  1.123

User Model versus Baseline Model:

Comparative Fit Index (CFI)               0.858         0.875
Tucker-Lewis Index (TLI)                 0.830         0.850

Robust Comparative Fit Index (CFI)        0.874
Robust Tucker-Lewis Index (TLI)          0.848

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)              -5556.937     -5556.937
Loglikelihood unrestricted model (H1)     -5452.548     -5452.548

Akaike (AIC)                              11185.875     11185.875
Bayesian (BIC)                             11312.791     11312.791
Sample-size adjusted Bayesian (BIC)       11198.667     11198.667

Root Mean Square Error of Approximation:

RMSEA                                     0.066         0.058
90 Percent confidence interval - lower    0.053         0.045
90 Percent confidence interval - upper    0.078         0.070
P-value RMSEA <= 0.05                    0.021         0.147
Robust RMSEA                               0.062
```

FONTE: ADAPTADO DE R CORE TEAM (2020) E ROSSEEL (2012).

Após a verificação dos pressupostos da AFC, passa-se ao estudo da identificação do modelo estrutural, ou seja, do instrumento conforme representado na **Figura 4**. Para tanto, ele deve atender às condições de ordem e ordenação, com a primeira satisfeita se GL é maior ou igual

a 0 (em acordo ao demonstrado anteriormente,  $GL = 100$ ) e a segunda, por exemplo, quando cada constructo tem três ou mais indicadores (Hair, Babin et al., 2005), também satisfeita, considerando que o QMPR apresenta quatro indicadores para cada um.

O desenvolvimento da AFC no R (R Core Team, 2020) demanda, ainda, a seleção do algoritmo para se estimar os parâmetros livres (Hair et al., 2010), tarefa que se embasa na distribuição dos dados que se usará na técnica. Da **Figura 6**, observou-se que estes não atendem à normalidade multivariada, impelindo à escolha de um dos seguintes métodos: Assintótico Isento de Pressupostos de Distribuição (ADF), WLSM, WLSMV, S-B *scaling* ou *Bootstrapping* (Finney & DiStefano, 2006; Pilati & Laros, 2007). Optou-se pela estimação por máxima verossimilhança com erros robustos e teste estatístico ajustado por Satorra-Bentler (S-B *scaling*), disponível no pacote “lavaan” a ser instalado no R (R Core Team, 2020; Rosseel, 2012). A **Figura 7** traz os comandos e resultados (parciais para fins de leioute).

A primeira área em destaque contém a especificação do modelo do QMPR, com os indicadores associados aos respectivos constructos, em acordo à representação na **Figura 4**. Na próxima área apontam-se as covariâncias não esperadas entre os constructos, ou seja, de HA com U e de HA com R (**Figura 5**). Os próximos campos contêm índices de ajuste do modelo (robustos), os quais mensuram a “qualidade” deste, e devem apresentar os valores recomendados na **Tabela 3**.

**TABELA 3**  
VALORES RECOMENDADOS DE ÍNDICES DE AJUSTE DOS MODELOS ESTIMADOS

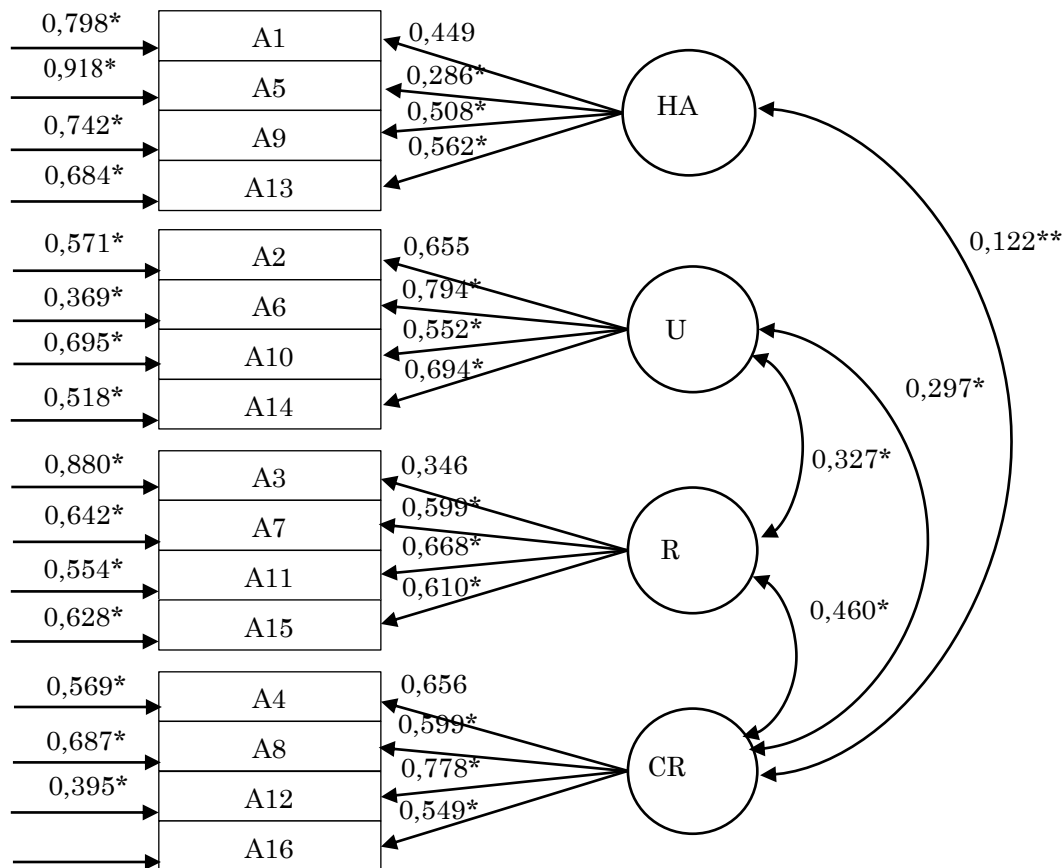
Índice	Valor
$\chi^2$	P-valores maiores do que 0,05 (significância 5%)
CFI	Igual ou maior do que 0,95
TLI	Maior do que 0,95
RMSEA	Menor do que 0,08

FONTE: ADAPTADO DE HAIR ET AL. (2010).

O qui-quadrado ( $\chi^2$ ), índice de ajuste comparativo (CFI), índice de Tucker Lewis (TLI) e o erro quadrático médio de aproximação (RMSEA) representam, nesta ordem, a diferença entre a matriz de covariância fornecida e a estimada; a comparação entre o modelo estimado e o nulo; esta comparação em se tratando de qui-quadrados normalizados; e a qualidade de ajuste se a estimação ocorresse na população (Hair, Babin et al., 2005; Hair et al., 2010). Como p-valor de  $\chi^2$  foi menor do 0,05 e CFI e TLI, respectivamente, tiveram valores de 0,875 e 0,850, considerou-se o modelo com “qualidade” inadequada.

FIGURA 8

MODELO PREVISTO NO QMPR E ESTIMATIVAS DAS CARGAS DE CONSTRUCTOS, COVARIÂNCIAS E ERROS



\*Significante a 95% de confiança.  
 \*\*Não-significante a 95% de confiança.

FONTE: ADAPTADO DE KEMBER ET AL. (2000), R CORE TEAM (2020) E ROSSEEL (2012).

O pacote lavaan (Rosseel, 2012) disponibiliza, também, as cargas dos constructos nos indicadores, as covariâncias estimadas entre aqueles e os erros de mensuração estimados, que figuram, nesta ordem, como setas dos constructos aos indicadores, setas entre os constructos e termos “ε” na **Figura 5**. Nos resultados expostos na **Figura 7**, eles constariam, em seguida, padronizados nas colunas “Std.all” (cargas no campo “Latent Variables”, covariâncias em “Covariances” e erros em “Variances”). Assim, conforme a **Figura 8**, as cargas nos indicadores se mostraram inadequadas, ou seja, menores do que 0,5 em seu valor padronizado (Hair et al., 2010), para HA em A4 (0,286) e R em A3 (0,346). Os erros, menores do que 2,5 (Hair et al., 2010), são aceitáveis. Os erros, cargas e covariâncias, em geral, foram significantes ao nível de 5%, exceto a covariância entre HA e CR. Os indicadores A1, A2, A3 e A4, padronizados de maneira automática no R, receberam o valor 1 (em acordo ao exposto



anteriormente), o que implicou ausência do cálculo de p-valor. Para se sobrepujar tal problema, pode-se atribuir o número a outros indicadores na definição do modelo (por exemplo, como a expressão “HA =~ A5 + A1 + A9 + A13” o faz para A5).

Por fim, destaca-se que o pacote lavaan (Rosseel, 2012) fornece índices de modificação (MI), que consistem no nível de melhoria de  $\chi^2$  para estimativas dos parâmetros que se encontravam fixos. Valores de MI acima de 4 implicam melhorias expressivas, embora devam estar embasadas no arcabouço teórico do instrumento (Hair et al., 2010). Conforme a **Figura 9** (mostrada parcialmente), há três tipos de estimativas no QMPR para cômputo de MI: covariâncias entre os constructos (HA e U; HA e R); cargas destes nos indicadores aos quais não estão relacionados no modelo; e covariâncias entre indicadores (nesta ordem, os tipos são indicados pelos operadores “~~”, “=~” e “~”). A melhoria mais importante ocorreria ao prever covariância entre HA e U (MI = 32,604) (**Figura 9**), ou seja, ao excluir na especificação do modelo (**Figura 7**) a expressão “HA~0\*U”. Em seguida, a AFC deveria ser empregada novamente para se avaliar a existência de aprimoramentos, desde que convergissem à teoria.

**FIGURA 9**  
ÍNDICES DE MODIFICAÇÃO PARA O QMPR

```
> modificationIndices(fit.modelo_estrutural)
```

	lhs	op	rhs	mi	epc	sepc.lv	sepc.all	sepc.nox
17	HA	~~	U	32.604	-0.168	-0.537	-0.537	-0.537
18	HA	~~	R	0.189	0.007	0.044	0.044	0.044
43	HA	=~	A2	0.131	0.051	0.025	0.026	0.026

FONTE: ADAPTADO DE R CORE TEAM (2020) E ROSSEEL (2012).

Ressalta-se, ainda, que a validade de conteúdo, assim como a de critério, é empreendida para se analisar quão adequados são os indicadores ao representar um constructo (Babbie, 2016; Fitzner, 2007). Ela pode ser apreciada por abordagem qualitativa e quantitativa (Souza et al., 2017), sendo que neste caso, as ferramentas indicadas para análise da validade de face são passíveis de utilização. Por consistir em uma revisão exaustiva do instrumento (Fitzner, 2007), as análises de conteúdo podem ocorrer imediatamente após a de face.

## CONSIDERAÇÕES FINAIS

Este artigo abordou a validação de instrumentos psicométricos por meio de métodos e técnicas empregadas no software R (R Core Team, 2020). Forneceu comandos e exemplificou resultados de análise de confiabilidade e validade, o que se concretizou pelo cômputo de



coeficientes, como o Alfa de Cronbach e ômega, pelo cálculo de índices de regressão e de concordância e por meio do emprego da análise fatorial confirmatória (AFC). A avaliação de pressupostos à AFC também foi realizada.

Os caminhos indicados e as discussões desenvolvidas aqui, a despeito de não consistirem em panaceia aos desafios envolvidos na validação de instrumentos psicométricos, contribuem à redução da complexidade da área, além de promover o acesso a uma ferramenta poderosa e gratuita a pesquisadores que se deparam com dificuldades em utilizar o software R.

## REFERÊNCIAS

- Babbie, E. (2016). *The practice of social research* (14th ed.). Boston: Cengage Learning.
- Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2014). *Testagem e avaliação psicológicas: introdução a testes e medidas* (8th ed.). Porto Alegre: AMGH.
- Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient Alpha and successor procedures. *Educational and Psychological Measurement*, 64, 391–418. <http://epm.sagepub.com/content/64/3/391>
- Delgado, R., & Tibau, X.-A. (2019). Why Cohen's Kappa should be avoided as performance measure in classification. *PLoS ONE*, 14(9), 1–26. <https://doi.org/10.1371/journal.pone.0222916>
- Finney, S. J., & DiStefano, C. (2006). Non-Normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269–314). Connecticut: Information Age Publishing.
- Fitzner, K. (2007). Reliability and validity: a quick review. *The Diabetes EDUCATOR*, 33(5), 775–780. <http://tde.sagepub.com/content/33/5/775>
- Gwet, K. L. (2002). Kappa Statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical Methods for Inter-Rater Reliability Assessment*, (1), 1–5. [https://www.researchgate.net/profile/Kilem\\_Gwet/publication/246598896\\_Kappa\\_Statistic\\_is\\_not\\_Satisfactory\\_for\\_Assessing\\_the\\_Extent\\_of\\_Agreement\\_Between\\_Raters/links/5ae333f7a6fdcc9139a18814/Kappa-Statistic-is-not-Satisfactory-for-Assessing-the-Extent-of-A](https://www.researchgate.net/profile/Kilem_Gwet/publication/246598896_Kappa_Statistic_is_not_Satisfactory_for_Assessing_the_Extent_of_Agreement_Between_Raters/links/5ae333f7a6fdcc9139a18814/Kappa-Statistic-is-not-Satisfactory-for-Assessing-the-Extent-of-A)
- Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, 61(1), 29–48.

- Gwet, K. L. (2019a). *Calculating chance-corrected agreement coefficients (CAC)*. <https://cran.r-project.org/web/packages/irrCAC/vignettes/overview.html>
- Gwet, K. L. (2019b). *irrCAC: computing chance-corrected agreement coefficients (CAC)*. <https://cran.r-project.org/package=irrCAC>
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2005). *Análise multivariada de dados* (5th ed.). Porto Alegre: Bookman.
- Hair, J. F. J., Babin, B., Money, A. H., & Samouel, P. (2005). *Fundamentos de métodos de pesquisa em Administração*. Porto Alegre: Bookman.
- Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Nova Jersey: Prentice Hall.
- Hardesty, D. M., & Bearden, W. O. (2004). The use of expert judges in scale development: implications for improving face validity of measures of unobservable constructs. *Journal of Business Research*, 57(2), 98–107. [https://doi.org/10.1016/S0148-2963\(01\)00295-8](https://doi.org/10.1016/S0148-2963(01)00295-8)
- Kelley, K. (2020). *MBESS: The MBESS R package*. <https://cran.r-project.org/package=MBESS>
- Kember, D., Leung, D. Y. P., Jones, A., Loke, A. Y., McKay, J., Sinclair, K., ... Yeung, E. (2000). Development of a questionnaire to measure the level of reflective thinking. *Assessment & Evaluation in Higher Education*, 25(4), 381–395. <https://doi.org/10.1080/713611442>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Nova Iorque: The Guilford Press.
- Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2), 151–162. <https://journal.r-project.org/archive/2014-2/korkmaz-goksuluk-zararsiz.pdf>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Lehman, A., O'Rourke, N., Hatcher, L., & Stepanski, E. J. (2013). *JMP for basic univariate and multivariate statistics: methods for researchers and social scientists* (2nd ed.). Cary: SAS Institute Inc.
- Mecklin, C. J., & Mundfrom, D. J. (2005). A Monte Carlo comparison of the Type I and Type II error rates of tests of multivariate normality. *Journal of Statistical Computation and Simulation*, 75(2), 93–107. <https://doi.org/10.1080/0094965042000193233>

- Melan, R. L. (2018). *Pensamento reflexivo para sustentabilidade: Considerações ao ensino de administradores e validação de um instrumento* (Dissertação de mestrado, Universidade Estadual de Londrina). <http://www.bibliotecadigital.uel.br/document/?code=vtls000218805>
- Pacico, J. C., & Hutz, C. S. (2015). Validade. In C. S. Hutz, D. R. Bandeira, & C. M. Trentini (Eds.), *Psicometria* (pp. 71–83). Porto Alegre: Artmed.
- Pasquali, L. (2009). Psicometria. *Revista da Escola de Enfermagem da USP*, 43(spe), 992–999. <https://doi.org/10.1590/S0080-62342009000500002>
- Pilati, R., & Laros, J. A. (2007). Modelos de equações estruturais em Psicologia: conceitos e aplicações. *Psicologia: Teoria e Pesquisa*, 23(2), 205–216. <https://doi.org/10.1590/S0102-37722007000200011>
- R Core Team. (2020). *R: a language and environment for statistical computing*. <https://www.r-project.org/>
- Rosseel, Y. (2012). lavaan: An R package for structural equation. *Journal of Statistical Software*, 48(2). <https://www.jstatsoft.org/article/view/v048i02>
- Santos, R. J. (2017). *Matrizes, vetores e geometria analítica*. <https://www.ime.unicamp.br/~deleo/MA141/ld01a.pdf>
- Souza, A. C. de, Alexandre, N. M. C., & Guirardello, E. de B. (2017). Propriedade psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. *Epidemiologia e Serviços de Saúde*, 26(3), 649–659. <https://doi.org/10.5123/s1679-49742017000300022>
- Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: a practical guide to their development and use* (15th ed.). Oxford: Oxford University Press.
- Teo, T., & Fan, X. (2013). Coefficient alpha and beyond: Issues and alternatives for educational research. *Asia-Pacific Education Researcher*, 22(2), 209–213. <https://doi.org/10.1007/s40299-013-0075-z>
- Valentini, F., & Damásio, B. F. (2016). Variância média extraída e confiabilidade composta: indicadores de precisão. *Psicologia: Teoria e Pesquisa*, 32(2), 1-7. <https://doi.org/10.1590/0102-3772e322225>.
- Zaichkowsky, J. L. (1985). Measuring the involvement construct. *Journal of Consumer Research*, 12(3), 341–352. <https://www.jstor.org/stable/254378>
- Zanon, C., & Hauck Filho, N. (2015). Fidedignidade. In C. S. Hutz, D. R. Bandeira, & C. M. Trentini (Eds.), *Psicometria* (pp. 84–94). Porto Alegre: Artmed.

---

**INFORMAÇÕES ACADÊMICAS E PROFISSIONAIS DA AUTORIA****THAIS ACCIOLY BACCARO**

Doutora em Administração de Organizações, Universidade de São Paulo, Brasil.  
Docente no Departamento de Administração da Universidade Estadual de Londrina, Brasil.  
Docente no Programa de Pós-Graduação em Administração da Universidade Estadual de Londrina, Brasil.  
ORCID iD: 0000-0001-6093-8429 | ID Lattes: 9552226887397427  
E-mail: thaisbaccaro@uel.br (autora correspondente)

**RODRIGO LIBANEZ MELAN**

Doutorando em Administração, Universidade Estadual de Maringá, Brasil.  
Mestre em Administração, Universidade Estadual de Londrina, Brasil.  
ORCID iD: 0000-0003-3880-7451 | ID Lattes: 5317599840640368  
E-mail: rodrigo.libanezm@gmail.com

**SAULO FABIANO AMÂNCIO VIEIRA**

Doutor em Administração, Universidade Nove de Julho, Brasil.  
Docente no Departamento de Administração da Universidade Estadual de Londrina, Brasil.  
Docente no Programa de Pós-Graduação em Administração da Universidade Estadual de Londrina, Brasil.  
ORCID iD: 0000-0003-0319-7392 | ID Lattes: 5527654035542164  
E-mail: saulo@uel.br

**CONTRIBUIÇÃO E ATUAÇÃO EM ATIVIDADES DE AUTORIA****THAIS ACCIOLY BACCARO**

Orientou atividades envolvidas com Definições / Conceitos / Teorias, Pesquisa / Seleção / Organização / Administração dos Dados, Exame Pormenorizado / Avaliação Crítica / Validação dos Dados e Descrição e Delimitação dos Procedimentos Metodológicos.  
Contribuiu e atuou igualmente, em relação à equipe de autoria, nas atividades envolvidas com Responsabilização na Pesquisa Documentada no Artigo Submetido, Escolha e Manuseio de Software e Redação do Texto.

**RODRIGO LIBANEZ MELAN**

Contribuiu e atuou igualmente, em relação à equipe de autoria, nas atividades envolvidas com Definições / Conceitos / Teorias, Pesquisa / Seleção / Organização / Administração dos Dados, Exame Pormenorizado / Avaliação Crítica / Validação dos Dados, Descrição e Delimitação dos Procedimentos Metodológicos, Responsabilização na Pesquisa Documentada no Artigo Submetido, Escolha e Manuseio de Software e Redação do Texto.

**SAULO FABIANO AMÂNCIO VIEIRA**

Orientou atividades envolvidas com Definições / Conceitos / Teorias e Pesquisa / Seleção / Organização / Administração dos Dados.  
Contribuiu e atuou igualmente, em relação à equipe de autoria, nas atividades envolvidas com Exame Pormenorizado / Avaliação Crítica / Validação dos Dados, Descrição e Delimitação dos Procedimentos Metodológicos, Responsabilização na Pesquisa Documentada no Artigo Submetido e Redação do Texto.  
Ofereceu apoio e assistência em atividades envolvidas com Escolha e Manuseio de Software.

**DECLARAÇÃO DE CONFLITOS DE INTERESSES DA AUTORIA****THAIS ACCIOLY BACCARO**

Declarou não possuir conflitos de interesse verdadeiro (factual), possível (potencial), de ordem financeira, de ordem pessoal, de ordem acadêmica, relacionado à afiliação institucional, de ordem política e de ordem religiosa. Além dos questionados, declarou não possuir outros conflitos de interesses.

**RODRIGO LIBANEZ MELAN**

Declarou não possuir conflitos de interesse verdadeiro (factual), possível (potencial), de ordem financeira, de ordem pessoal, de ordem acadêmica, relacionado à afiliação institucional, de ordem política e de ordem religiosa. Além dos questionados, declarou não possuir outros conflitos de interesses.

**SAULO FABIANO AMÂNCIO VIEIRA**

Declarou não possuir conflitos de interesse verdadeiro (factual), possível (potencial), de ordem financeira, de ordem pessoal, de ordem acadêmica, relacionado à afiliação institucional, de ordem política e de ordem religiosa. Além dos questionados, declarou não possuir outros conflitos de interesses.