

 <p>ESCOLA DE CIÊNCIAS DA SAÚDE E DA VIDA</p>	<p><b>PSICO</b></p> <p>Psico, Porto Alegre, v. 55, n. 1, p. 1-12, jan.-dez. 2024 e-ISSN: 1980-8623   ISSN-L: 0103-5371</p>
<p><a href="http://dx.doi.org/10.15448/1980-8623.2024.1.40882">http://dx.doi.org/10.15448/1980-8623.2024.1.40882</a></p>	

SEÇÃO: ARTIGOS

## Passos para a elaboração de instrumentos de escolha forçada do tipo rank

*Steps for the construction of rank type forced choice instruments*

*Pasos para la elaboración de instrumentos en el formato rank de elección forzada*

**Paloma Pereira de Almeida<sup>1</sup>**

[orcid.org/0000-0002-9291-9753](https://orcid.org/0000-0002-9291-9753)  
[palomapalmeida@gmail.com](mailto:palomapalmeida@gmail.com)

**Juliane Callegaro**

**Borsa<sup>2</sup>**

[orcid.org/0000-0001-7703-5509](https://orcid.org/0000-0001-7703-5509)  
[juliborsa@gmail.com](mailto:juliborsa@gmail.com)

**J. Landeira-Fernandez<sup>3</sup>**

[orcid.org/0000-0002-8395-8008](https://orcid.org/0000-0002-8395-8008)  
[landeira@puc-rio.br](mailto:landeira@puc-rio.br)

**Recebido em:** 3 maio 2021.

**Aprovado em:** 3 dez. 2021.

**Publicado em:** 22 out. 2024.

**Resumo:** Os instrumentos no formato de escolha forçada têm se mostrado bastante úteis na prevenção do falseamento das respostas. No entanto, mesmo sabendo que a vantagem do controle deste viés é bastante útil em diversos contextos de atuação do psicólogo, o número de instrumentos psicológicos neste formato e artigos brasileiros sobre esta temática ainda é escasso. O presente artigo tem como objetivo fomentar esta discussão, trazendo informações sobre as principais etapas presentes na construção de um instrumento de escolha forçada do tipo RANK, com vistas a auxiliar os pesquisadores em sua execução. As etapas da criação dos blocos de respostas e a forma de análise da deseabilidade social dos itens são alguns dos tópicos discutidos.

**Palavras-chave:** deseabilidade social, psicometria, validade dos testes

**Abstract:** Instruments in the forced choice format have been shown to be very useful in preventing the falsification of responses. However, even knowing that the advantage of controlling this bias is quite useful in several contexts of the psychologist's performance, the number of psychological instruments in this format and Brazilian articles on this theme is still scarce. The purpose of the present article is to foster this discussion, providing information on the main stages in the construction of an instrument of forced choice of the RANK type scale in order to assist researchers in the process of execution of this scale. The paper discusses some topics regarding the stages of the creation of the blocks and the way of analysing the social desirability of the items.

**Keywords:** social desirability, psychometrics, validity of tests

**Resumen:** Se ha demostrado que los instrumentos en el formato de elección forzada son muy útiles para prevenir la falsificación de las respuestas. Sin embargo, mismo sabiendo que hay ventaja de controlar este sesgo y esto es bastante útil en varios contextos de la actuación del psicólogo, el número de instrumentos psicológicos en este formato y artículos brasileños sobre este tema es aún escaso. Este artículo tiene como objetivo fomentar esta discusión, aportando información sobre las principales etapas en la construcción de un instrumento de elección forzada del tipo RANK, con el objetivo de ayudar a los investigadores en sus trabajos. Las etapas de la creación de los bloques y la forma de analizar la deseabilidad social de los ítems son algunos de los temas presentados.

**Palabras-clave:** deseabilidad social, psicometria, validez



Artigo está licenciado sob forma de uma licença  
[Creative Commons Atribuição 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/).

Os instrumentos no formato de escolha forçada têm se mostrado bastante úteis para a prevenção do falseamento das respostas (Brown, 2014; Brown & Maydeu-Olivares, 2011, 2013, 2018; Cao & Drasgow, 2019;

<sup>1</sup> Pesquisadora autônoma, Rio de Janeiro, RJ, Brasil.

<sup>2</sup> Pesquisadora autônoma, Porto Alegre, RS, Brasil.

<sup>3</sup> Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, RJ, Brasil.

Carvalho & Ambiel, 2017; Roger-Welter & Capitão, 2007). Diferente do formato Likert, que permite ao sujeito a possibilidade de atribuir uma pontuação alta ou média a todos os itens ou até mesmo assinalar a opção neutra da escala (quando existe), diminuindo, assim, a transparência em suas respostas (Carvalho & Ambiel, 2017; Roger-Welter & Capitão, 2007), o formato da escolha forçada faz com que os participantes sejam forçados a atribuir um valor diferente aos itens, sem a possibilidade de concordar com todos eles, o que gera uma menor distorção decorrente da desejabilidade social (Brown, 2014; Brown & Maydeu-Olivares, 2011, 2013, 2018). Afinal, o ato do indivíduo ter que ranquear itens igualmente desejáveis, simula uma situação semelhante ao processo de escolha real que o indivíduo tem que fazer na vida, no qual é forçado a escolher, por exemplo, qual formação acadêmica irá seguir ou qual emprego vai aceitar, simulando, assim, sua escolha diante de comportamentos que lhe são mais característicos (Brown, 2014; Brown & Maydeu-Olivares, 2011, 2013; Meade, 2004; Roger-Welter & Capitão, 2007). Entende-se, portanto, que para realizar tais escolhas no questionário, o sujeito precisará buscar referências internas, olhando para dentro de si. Ao aplicar um valor diferente para cada item, elimina-se a possibilidade de concordar apenas com os itens de conteúdo desejável amenizando, assim, o viés da desejabilidade social (Roger-Welter & Capitão, 2007).

Por desejabilidade social, entende-se que é um viés referente a demandas culturais e/ou pessoais que comprometem a resposta do sujeito mediante a situação de testagem, falseando, assim, a sua resposta. Algumas vezes, este falseamento é causado de forma proposital com o objetivo de parecer melhor para impressionar o avaliador, mas também pode ser apenas consequência de uma falta de autoconhecimento ou de dificuldade em aceitar suas próprias limitações não caracterizando, desta forma, um falseamento consciente (Carvalho & Ambiel, 2017; Roger-Welter & Capitão, 2007).

O formato da escolha forçada permite algumas formas de organização dos itens. No formato

PICK o respondente deve escolher o item que mais lhe caracteriza, selecionando apenas um item preferido em meio a um bloco. No formato MOLE (acrônimo para MOst and LEast), o sujeito deve indicar os itens que mais e menos se assemelham com ele. Ainda, é possível o formato do preenchimento completo, também chamado de RANK, no qual o respondente terá que ranquear todos os itens apresentados em uma ordem de preferência, de forma hierarquizada, sem deixar nenhum em branco (Brown & Maydeu-Olivares, 2018; Cao & Drasgow, 2019).

Tais formatos requerem o uso de medidas ipsativas. O termo ipsativo é usado de forma genérica como sinônimo para "dados interdependentes" e tal interdependência gerada altera as propriedades psicométricas da escala pois o item selecionado não depende só do nível de latência do traço que está sendo medido pelo item, mas também do conjunto de itens do qual faz parte, fazendo com que cada escore observado seja contaminado pelos demais escores do conjunto de itens. Ou seja, do ponto de vista matemático, os dados são considerados ipsativos quando a soma dos escores de um questionário é igual para todos os respondentes. Desta forma, os dados ipsativos comprometem importantes pressupostos da teoria clássica dos testes, como, por exemplo, a independência da variância do erro e o critério paramétrico padronizado pois o nível de medida não é ordinal e nem escalar. Porém, mesmo diante de tais complexidades provenientes de uma psicometria avançada, pesquisadores afirmam que as vantagens que envolvem este tipo de instrumento superam suas dificuldades de análises estatísticas, sobretudo quando há risco de falseamento de resposta devido ao viés da desejabilidade social (Brown, 2014; Brown & Maydeu-Olivares, 2011, 2013, 2018; Roger-Welter & Capitão, 2007).

Um modelo teórico que abarca esta forma de avaliação e que é utilizado como grande referência para os instrumentos de escolha forçada seria o proposto por Thurstone (1931), chamado de Teoria de Resposta ao Item Thurstoniana (TRI-T) e, posteriormente, adaptado por Brown &

Maydeu-Olivares (2011, 2013, 2018). Neste modelo, um dos principais conceitos a ser considerado é o da "utilidade". Trata-se do afeto envolvido pelo respondente na escolha de determinado item, ou seja, é a própria motivação do sujeito que o leva a preferir determinado item em detrimento de outro. Assim, entende-se que os itens mais "úteis" são aqueles preferidos pelo respondente. Tal "utilidade" não pode ser observada diretamente, pois se trata de uma variável explicada por meio de modelos latentes.

No entanto, mesmo sabendo que a vantagem do controle do viés da desejabilidade social é bastante útil em diversos contextos de atuação do psicólogo (escola, empresa, clínica, entre outros), o número de instrumentos psicológicos neste formato ainda é escasso. Existem alguns testes neste formato disponíveis no Sistema de Avaliação de Testes Psicológicos (SATEPSI), como o HumanGuide (Roger-Welter, 2014) e o Quati (Zacharias, 2000), porém ainda são poucas as opções dentre o universo existente de testes disponíveis. Da mesma forma, há carência de artigos sobre medidas ipsativas e instrumentos de escolha forçada no Brasil. Semelhante à busca que os autores Roger-Welter & Capitão (2007) realizaram, foi feito pelos autores deste manuscrito um rápido levantamento em dezembro de 2021 nas bases Scientific Electronic Library Online (SciELO), Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS) e Biblioteca Virtual em Saúde (BVS) utilizando as palavras-chave "ipsatividade", "medida ipsativa", "ipsativos" e "escolha forçada" com o filtro para os artigos em português. Ao final, foram encontrados apenas dois manuscritos que abordavam a temática das medidas ipsativas (Roger-Welter & Capitão, 2007; Valentini, 2018). Considerando esses aspectos, o presente artigo tem como objetivo fomentar esta discussão, trazendo informações sobre as principais etapas presentes na construção de um instrumento de escolha forçada do tipo RANK, com vistas a auxiliar os pesquisadores na sua execução.

## Etapas do processo de construção de instrumentos de escolha forçada do tipo RANK

### *Etapa 1: definição do construto*

Esta é uma etapa muito importante na construção de um instrumento psicológico, independentemente de ser no formato da escolha forçada ou não, pois a conceituação do construto a ser avaliado deve servir como guia norteador para a ferramenta a ser construída e, sendo assim, está diretamente relacionado com a garantia de evidências de validade de conteúdo. Posto isto, faz-se necessário elaborar as definições: (a) constitutiva, que descreve conceitualmente o construto para o qual o instrumento será construído a partir da teoria que o embasa; e (b) operacional, que deve ser realizada em termos de operações concretas, ou seja, é necessário descrever comportamentos físicos por meio dos quais o construto se expressa. Em outras palavras, trata-se de definir o construto em termos de comportamentos observáveis organizados em categorias de comportamento que serão a base para a elaboração dos itens. É fundamental que esta definição reflita o traço latente com a máxima semelhança possível (Borsa & Size, 2017; Carvalho & Ambiel, 2017).

### *Etapa 2: escrita dos itens da escala*

Após a devida conceituação constitutiva e operacional do construto, iniciam-se os esforços para a elaboração dos itens do instrumento. Esta etapa é crucial pois, dependendo do modo como determinado item é elaborado, ele pode acabar avaliando construtos diferentes do que se pretende (Borsa & Size, 2017). Para elaborá-los, é fundamental atentar para alguns critérios: (a) comportamental – o item deve, necessariamente, expressar um comportamento; (b) desejabilidade – o item deve abordar comportamentos característicos para cada construto; (c) simplicidade – deve expressar uma única ideia; (d) clareza – deve ser inteligível para a população-alvo; e (e) relevância – deve ser consistente com o traço

definido (AERA et al., 2014).

Recomenda-se que o total de itens elaborados seja bem maior que a quantidade que se espera na versão final. Esta estimativa depende do fenômeno que se pretende mensurar, porém é consenso a importância de elaborar no momento inicial uma quantidade maior de itens, uma vez que se sabe que alguns itens são excluídos após a análise dos juízes especialistas e leigos, e que as análises estatísticas revelarão outros como não indicados (Clark & Watson, 2019).

Algo importante a ser considerado no contexto dos instrumentos de escolha forçada do tipo RANK, é que é preferível que os itens sejam escritos no sentido positivo e negativo de modo que possam abarcar todo o contínuo do traço: do polo baixo ao polo mais alto do construto (Brown, 2014; Brown & Maydeu-Olivares, 2011, 2013, 2018). Por exemplo, um item considerado positivo poderia ser "eu sou um líder extrovertido" e um negativo seria "eu sou um líder introvertido".

### *Etapa 3: busca de evidências de validade de conteúdo – análise dos itens*

Após a escrita dos itens, faz-se necessário certificar-se que aqueles selecionados para compor a versão preliminar do instrumento de escolha forçada a ser construído possuem as devidas evidências de validade de conteúdo. Para tanto, são necessários dois grupos distintos de juízes, os especialistas e os leigos, conforme veremos a seguir.

**Juízes especialistas.** Os juízes especialistas devem ser pessoas com alta expertise na área (AERA et al., 2014). Para analisar o grau de concordância dos itens, existem diversas opções encontradas na literatura. Citando algumas, temos o Agreement Coefficient<sup>1</sup> (AC1) (Gwet, 2008), o Coeficiente de Validade de Conteúdo do Item (CVC) (Hernández-Nieto, 2002), o Índice de Validade de Conteúdo do Item (i-IVC) (Polit & Beck, 2006), o Índice Kappa (k) (Landis e Koch, 1977) e o Coeficiente de Correlação Intraclasse (CCI) (Shrout & Fleiss, 1979). Como exemplos, a fim de explicar com mais detalhes sobre este processo ao leitor, serão apresentados os conceitos dos dois últimos.

O Índice Kappa (k) é a razão da proporção de vezes que os juízes especialistas concordam (corrigido por concordância devida ao acaso) com a proporção máxima de vezes que deveriam concordar (corrigido por concordância devida ao acaso). Vale salientar que esse índice é aplicável apenas quando os dados são categóricos e estão em uma escala nominal (Alexandre & Coluci, 2011; Borsa & Size, 2017; Siegel & Castellan, 2006). Para a realização da análise do seu resultado, sugere-se utilizar a classificação de Landis e Koch (1977):

- Kappa < 0 = sem concordância.
- Kappa entre 0 e 0,19 = concordância pobre.
- Kappa entre 0,20 e 0,39 = concordância baixa.
- Kappa entre 0,40 e 0,59 = concordância moderada.
- Kappa entre 0,60 e 0,79 = concordância substancial.
- Kappa entre 0,80 e 1,00 = concordância quase perfeita.

O Coeficiente de Correlação Intraclasse (CCI) é indicado para mensurar a homogeneidade de duas ou mais medidas e, assim, pode ser utilizado para medir a concordância entre especialistas. Para uma adequada interpretação dos valores do CCI, sugere-se a classificação de Fleiss (1981):

- CCI < 0,40 = pobre;
- $0,40 \leq \text{CCI} < 0,75$  = satisfatória;
- CCI  $\geq 0,75$  = excelente.

É comum existir a necessidade de exclusão de alguns itens que não tiveram o grau de concordância satisfatório. Assim, os itens que passarem por essa etapa são os que deverão ser utilizados na próxima.

**Juízes leigos.** Trata-se do público-alvo do instrumento a ser construído. Os itens analisados nesta etapa devem ser apenas os itens que obtiveram uma boa avaliação dos juízes especialistas, de acordo com o cálculo escolhido. Os juízes leigos deverão ser responsáveis por dois tipos de análises (a) a semântica e (b) da desejabilidade social. Na avaliação semântica, o objetivo

é checar se o público-alvo compreende claramente os itens. Todos os itens que não estiverem devidamente claros deverão ser modificados e analisados novamente por outras pessoas ou eliminados. Pode-se, por exemplo, apresentar os itens aos participantes e, então, solicitar que ele/ela registre a sua compreensão através de uma escala Likert de cinco pontos, sendo 1 (não compreendi nada da afirmação) e 5 (compreendi a afirmação totalmente). Para obter essa informação acerca da clareza dos itens, calcula-se a média das respostas de cada um. Neste caso, os itens que obtiverem uma média acima de 4 podem ser considerados como aprovados pelos participantes.

Uma sugestão de avaliação da desejabilidade social, é solicitar que os participantes informem a sua percepção dos itens no que se refere ao seu grau de desejabilidade social utilizando, por exemplo, uma escala tipo Likert de nove pontos, sendo 1 (totalmente indesejável para mim) e 9 (totalmente desejável para mim). Para saber o valor, calcula-se a média de pontos que cada item obteve. Esta informação é importante para classificar os itens de acordo com sua desejabilidade social e facilitar a sua ordenação posterior em cada bloco (Brown & Maydeu-Olivares, 2011, 2013, 2018). Esta análise não envolve exclusão de itens, visto que o objetivo é somente classificá-los de acordo com o nível de sua desejabilidade. Algumas perguntas de estímulo podem ser feitas, como, "o quão essa frase é socialmente desejável para você?" ou "o quanto você gostaria de ter esse comportamento?", a fim de elucidar melhor ao participante o conceito de desejabilidade social.

Após essa avaliação, cada item terá o seu valor de desejabilidade social. Isto é, seguindo o exemplo da escala Likert de 9 pontos, os itens terão baixa (mais próximo do valor 1), média ou alta desejabilidade (mais próximo do valor 9). Tal informação serve somente para nortear a organização posterior dos blocos de respostas. A orientação é que cada bloco seja composto por itens com valores semelhantes, de modo que não haja em um mesmo bloco itens com valores muito discrepantes. Assim, os valores de corte para classificar os itens exatamente em

baixa, média ou alta desejabilidade não é o mais importante. Com esta classificação e organização, pode-se avançar para a próxima etapa.

#### *Etapa 4: organização dos itens para compor os blocos de escolha forçada*

Considerando que se trata de itens que irão compor um instrumento no formato de escolha forçada, a organização destes em blocos de respostas possui algumas particularidades se comparado a um instrumento de escala Likert, como, por exemplo, o fato de que escalas de escolha forçada não podem ser analisadas pela Teoria Clássica dos Testes, por resultarem em dados ipsativos como já mencionado anteriormente (Roger-Welter & Capitão, 2007). Desta forma, sugere-se adotar para tal organização o modelo da Teoria de Resposta ao Item Thurstoniana – TRI-T (Brown, 2014; Brown & Maydeu-Olivares, 2011, 2013, 2018). Tal modelo engloba, além das análises já descritas anteriormente, duas análises adicionais: a análise fatorial exploratória (AFE) e o agrupamento final dos itens. Assim, o objetivo principal desta etapa é o de construir blocos de itens que sejam adequados ao formato da escolha forçada do tipo RANK, de acordo com a TRI-T.

O modelo fatorial thurstoniano é um modelo de segunda ordem, no qual as utilidades latentes são fatores de primeira ordem e os traços latentes são fatores de segunda ordem. A diferença entre as utilidades latentes de dois itens  $\{i, k\}$  prediz a resposta a uma comparação binária  $\mathcal{Y}_{(i,k)}$ , conforme a equação (1).

$$\mathcal{Y}_{(i,k)} = \zeta_i - \zeta_k \quad (1)$$

No caso em que diferença de utilidades é positiva ( $> 0$ ), a dicotomização da comparação entre as duas variáveis é codificada como 1. Caso seja negativa ( $< 0$ ), é codificada como 0. Por sua vez, as utilidades latentes são preditas pelos traços latentes. Este modelo pode ser representado em notação matricial pela equação (2).

$$t = m + \Lambda\eta + \varepsilon \quad (2)$$

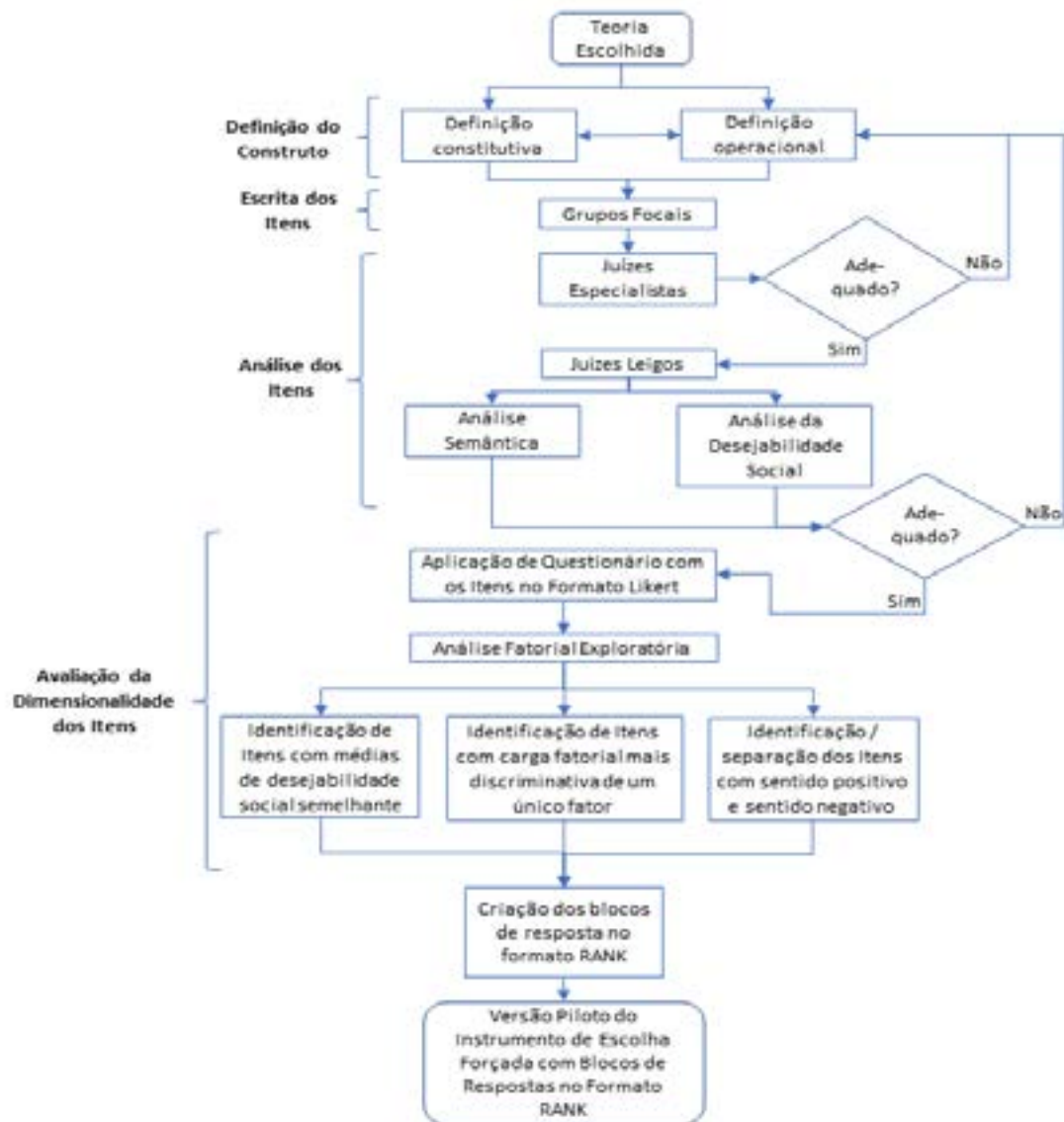
Nesta expressão,  $\mathbf{t}$  é um vetor de utilidades,  $\mathbf{m}$  é um vetor de interceptos dos itens,  $\Lambda$  é uma matriz de cargas fatoriais,  $\eta$  é um vetor de traços latentes,  $\epsilon$  é um vetor de erro de medida não correlacionado com os traços latentes (Brown & Maydeu-Olivares, 2018). Uma vez que o modelo thurstoniano é um modelo restrito (i.e., deve-se especificar quais cargas fatoriais serão estimadas e quais serão fixas em 0) temos um modelo fatorial confirmatório.

A AFE é amplamente utilizada nas pesquisas em psicologia. Trata-se de um conjunto de técnicas estatísticas que visam encontrar a estrutura subjacente em uma matriz de dados para, então, determinar a quantidade e os tipos das variáveis latentes, ou seja, os fatores que melhor representam os itens. Em outras palavras, este método reduz os dados provenientes de um conjunto de escores a um número menor de fatores (Damásio & Dutra, 2017). Para a realização da AFE, deve-se selecionar um grupo de pessoas do público-alvo do instrumento e aplicar neste grupo uma escala em formato Likert composta pelos itens avaliados de forma satisfatória pelos juízes especialistas e leigos. Assim, sabendo que os fatores são formados por um conjunto de itens que apresentam correlação entre si, o objetivo da aplicação de tal escala é o de posteriormente selecionar itens mais discriminativos de cada fator, para diminuir a correlação entre os itens de um mesmo bloco, pois blocos com conteúdos heterogêneos são mais indicados por aumentarem a quantidade de informação obtida para cada traço (Brown, 2014; Brown & Maydeu-Olivares, 2011, 2013, 2018). O propósito de realizar a AFE nesta etapa é o de explorar a estrutura fatorial dos dados para, subsequentemente, testar um modelo confirmatório com os itens de escolha forçada. Como a natureza dos dados provenientes de escalas de escolha forçada não permitem que modelos saturados sejam testados, os itens serão primeiramente testados em formato Likert para que a dimensionalidade deles seja avaliada (Brown & Maydeu-Olivares, 2018).

A partir dos resultados obtidos na AFE, já é possível formar os blocos de respostas do instru-

mento. Cada bloco deverá possuir as seguintes características: (a) ser composto por itens com médias de desejabilidade social semelhante; (b) ter, preferencialmente, metade dos itens no sentido positivo e metade no negativo, e (c) ser composto por itens com carga fatorial mais discriminativa de um único fator, ou seja, na perspectiva da TRI-T, cada item deve medir um único fator (i.e., a carga ser a mais discriminativa de um único traço). Deste modo, um item que carregue em dois ou três fatores, por exemplo, não deve ser prioridade ao construir um questionário de escolha forçada do tipo RANK (Brown & Maydeu-Olivares, 2018).

A Figura 1 apresenta um esquema metodológico proposto pelos autores, com base na TRI-T, para a construção de um instrumento de escolha forçada no formato RANK. Em algumas situações, pode haver mudanças nos passos propostos. Por exemplo, a utilização dos grupos focais para a elaboração dos itens não seria algo obrigatório, embora seja um passo importante para a garantia de bons itens. Considera-se que o conjunto de tais passos apresenta um caminho possível e metodologicamente adequado para a construção de tais instrumentos.



**Figura 1.** Visão geral das principais etapas para a elaboração de instrumentos no formato de escolha forçada do tipo RANK.

*Etapa 5: Busca de evidências de validade baseadas na estrutura interna – análise dos dados*

Considerando que nesta etapa a versão piloto do instrumento de escolha forçada está finalizada, o próximo passo é a aplicação do instrumento para uma amostra do seu público-alvo. Tendo em vista que o questionário de escolha forçada do tipo RANK possui o formato de full ranking, que é quando o respondente deve classificar as afirmações em uma ordem numérica de forma a não deixar nenhum item em branco, os dados

deverão apresentar propriedades ipsativas, caso os escores das pontuações sejam ponderadas como na Teoria Clássica dos Testes. Desta forma, as análises de dados que serão realizadas serão baseadas no método da TRI-T que modela a diferença das utilidades (o valor psicológico das preferências) relacionando-as com as escolhas observadas.

No formato de escolha forçada, os itens são comparados par-a-par, gerando variáveis dicotômicas para cada comparação. Nesse sentido, em uma *triplet* (conjunto de três itens por bloco)

$\{i, k, j\}$  temos três comparações binárias:  $\{i, k\}$ ,  $\{i, j\}$ ,  $\{k, j\}$ . Por sua vez, em uma *tetrad* (conjunto de quatro itens por blocos), temos seis comparações binárias. O número de comparações binárias em um bloco é dado pela expressão  $\tilde{n} = n(n-1)/2$ . As variáveis dicotômicas resultantes de tais comparações são codificados com base na utilidade do item, conceito proposto por Thurstone (1927). A utilidade é o valor psicológico atribuído ao item. Ou seja, o afeto psicológico que o item elicia. Dessa forma, em uma comparação binária  $\{i, k\}$ , o item  $i$  é preferido em relação ao item  $k$  quando o item  $i$  possui uma maior utilidade do que o item  $k$ . Por conseguinte, codificaríamos esta comparação binária como 1. Quando o segundo item da comparação é preferido em relação ao primeiro – isto é, o item  $k$  é preferido em relação ao item  $i$ , codificamos tal variável como 0. A utilidade é uma variável latente, portanto, não observada. O modelo de TRI pressupõe que a diferença de utilidades em cada comparação prediz a resposta observada.

Em seguida, sugere-se ajustar os dados a um *software* adequado para a realização dos cálculos estatísticos no modelo da TRI-T para dados ranqueados (Brown & Maydeu-Olivares, 2011, 2012). Como exemplos, pode-se utilizar o RStudio, que é gratuito, ou o Mplus. Sugere-se também o uso do estimador unweighted least squares mean and variance (ULSMV), conforme recomendado por Brown e Maydeu-Olivares (2011), por se tratar de dados categóricos. A sintaxe do modelo com a estrutura fatorial hipotetizada poderá ser gerada por meio de uma Macro do Excel disponibilizada por Brown e Maydeu-Olivares (2012). A qualidade do modelo deverá ser avaliada pelas cargas fatoriais, variâncias residuais e índices de ajuste. Recomenda-se que os valores de corte adotados sejam: Comparative Fit Index (CFI) > 0.95; Tucker-Lewis Index (TLI) > 0.95; Root Mean Square Error of Approximation (RMSEA) < 0.05 (Sivo, Fan, Witta & Willse, 2006).

Um exemplo de sintaxe de um modelo fatorial de segunda ordem, com três fatores, para o formato de escolha forçada no Mplus é apresentado a seguir. Breves explicações sobre a sintaxe

encontram-se de forma sublinhada, para melhor esclarecimento ao leitor.

TITLE: Exemplo de modelo fatorial de segunda ordem, com três fatores, para o formato de escolha forçada no Mplus;

DATA: FILE IS 'exemplo.txt';

VARIABLE: Nomeando todas as contidas no banco. Por exemplo, a comparação do item um com o item 2 está nomeada como "i1i2", ao passo que a comparação do item 1 com o item 3 foi nomeada como "i1i3". Neste exemplo, temos um banco contendo somente comparações binárias provenientes de um questionário de escolha forçada.

Names ARE

i1i2  
i1i3  
i2i3  
i4i5  
i4i6  
i5i6  
i7i8  
i7i9  
i8i9  
i10i11  
i10i12  
i11i12  
i13i14  
i13i15  
i14i15  
i16i17  
i16i18  
i17i18;

USEVARIABLES ARE i1i2-i17i18; Nomeando as variáveis que serão utilizadas na análise.

CATEGORICAL ARE ALL; Comando que define todas as variáveis como categóricas.

ANALYSIS:

ESTIMATOR = ulsmv; Estimador unweighted least squares com erros padrões robustos.

PARAMETERIZATION = theta;



MODEL: t14\*-1  
 t17\*1;

As variáveis binárias são preditas pelas utilidades ("t") dos itens.

t1 BY i1i2@1; t2 BY i1i2@-1;  
 t1 BY i1i3@1; t3 BY i1i3@-1;  
 t2 BY i2i3@1; t3 BY i2i3@-1;  
 t4 BY i4i5@1; t5 BY i4i5@-1;  
 t4 BY i4i6@1; t6 BY i4i6@-1;  
 t5 BY i5i6@1; t6 BY i5i6@-1;  
 t7 BY i7i8@1; t8 BY i7i8@-1;  
 t7 BY i7i9@1; t9 BY i7i9@-1;  
 t8 BY i8i9@1; t9 BY i8i9@-1;  
 t10 BY i10i11@1; t11 BY i10i11@-1;  
 t10 BY i10i12@1; t12 BY i10i12@-1;  
 t11 BY i11i12@1; t12 BY i11i12@-1;  
 t13 BY i13i14@1; t14 BY i13i14@-1;  
 t13 BY i13i15@1; t15 BY i13i15@-1;  
 t14 BY i14i15@1; t15 BY i14i15@-1;  
 t16 BY i16i17@1; t17 BY i16i17@-1;  
 t16 BY i16i18@1; t18 BY i16i18@-1;  
 t17 BY i17i18@1; t18 BY i17i18@-1;

Para dados ranqueados (tal como em triplets) o modelo assume que os erros das variáveis binárias possuem média 0.

i1i2-i17i18@0;

As utilidades são preditas pelos traços latentes. Neste exemplo, existem três traços latentes.

Valores com prior negativo (e.g., t1\*-1) são exemplos de itens com carga fatorial negativa.

Trait1 BY  
 t1\*-1  
 t4\*1  
 t7\*1  
 t10\*1  
 t13\*1  
 t16\*-1;

Trait2 BY  
 t2\*1  
 t5\*-1  
 t8\*1  
 t11\*1

Trait3 BY  
 t3\*1  
 t6\*1  
 t9\*-1  
 t12\*-1  
 t15\*1  
 t18\*1;

Para identificação do modelo, a variância dos fatores é fixada em 1.

Trait1-Trait3@1;

Colocando priors (valores iniciais) nas correlações. Ajuda a estimar e convergir o modelo com maior rapidez.

Trait1 WITH Trait2\*0 Trait3\*0;  
 Trait2 WITH Trait3\*0;

Fixa-se a variância única (uniqueness) de uma utilidade latente por bloco, para fins de identificação do modelo.

t1@1;  
 t4@1;  
 t7@1;  
 t10@1;  
 t13@1;  
 t16@1;

A precisão dos escores fatoriais deverá ser calculada com base na correlação entre o theta verdadeiro e o theta estimado ao quadrado (Brown & Maydeu-Olivares, 2011). Para realizar o cálculo da precisão, primeiramente simulamos os bancos de dados com base nas equações do modelo thurstoniano de escolha forçada. Assim, para 1.000 replicações, simulamos 1.000 bancos de dados que podem ser analisados posteriormente no Mplus em interface com R (R Core Team, 2019) por meio do pacote Mplus Automation (Hallquist & Willey, 2018). Os parâmetros de thresholds e as cargas fatoriais são simulados de acordo com os parâmetros obtidos no modelo testado empiricamente. O *theta* (ou escore fatorial) verdadeiro

é simulado pela função *mvrnorm* do R, no qual utilizamos: (1)  $\mu = 0$ , as correlações entre os fatores obtidas no modelo empírico, (3) o  $\eta$  da amostra. O tamanho da amostra na simulação é dado pelo tamanho da amostra empírica. Assim, no caso de a amostra testada empiricamente possuir 600 sujeitos, simulamos escores fatoriais para 600 observações. Estes parâmetros são inseridos na equação (3), a qual expressa a preferência do item  $j$  pelo item  $k$  na comparação  $l_i, k$ , caso em que codificamos a variável binária como 1.

$$-\gamma_{(l_i)} + (\lambda_i \eta_i - \lambda_k \eta_k) + (\varepsilon_i - \varepsilon_k) \geq 0 \quad (3)$$

Nesta expressão,  $\gamma$  representa o threshold,  $\lambda$  a carga fatorial,  $\varepsilon$  o erro. Assim, no caso em que a resposta do sujeito  $j$  para a comparação binária  $l_i, k$  possuir um valor maior ou igual a 0 na equação (3), a resposta simulada é codificada como 1. Caso contrário, é codificada como 0.

Em seguida, analisamos em 1.000 replicações no Mplus os bancos de dados simulados, os quais possuem as variáveis dicotomizadas de acordo com a equação (3). O modelo fatorial utilizado na análise é o mesmo testado na amostra empírica. Em cada uma das replicações, correlacionamos o *theta* verdadeiro ( $\eta_j$ ), o qual foi gerado pela função pela função *mvrnorm*, com o *theta* estimado ( $\hat{\eta}_j$ ) a partir dos bancos analisados no Mplus. Para tal cálculo da precisão, utilizamos a equação (4). Sendo assim, elevamos ao quadrado a correlação entre o *theta* verdadeiro e o *theta* estimado para obter a precisão dos escores.

$$\text{corr}(\eta_j, \hat{\eta}_j) = \sqrt{\rho} \quad (4)$$

### Etapa 6: busca de evidências de validade baseadas na relação com outras variáveis – análise dos dados

Para a busca de evidências de validade com outras variáveis, orienta-se que seja realizada uma Análise Fatorial Confirmatória (AFC) com os dados do instrumento escolhido para a comparação com o instrumento em construção. A AFC busca explicar as correlações entre as variáveis

observadas por meio das variáveis latentes subjacentes que se acredita predizê-las (Franco et al., 2017). O estimador recomendando seria o *weighted least squares mean and variance (WLSMV)*. Os dados do instrumento de escolha forçada a ser desenvolvido deverão ser analisados por meio da TRI-T (Brown & Maydeu-Olivares, 2011), adequado para dados de natureza ranqueada e estimador *unweighted least squares mean and variance (ULSMV)*. Ambas as análises se encontram no contexto de modelagem de equações estruturais e, para isso, sugere-se que sejam realizadas por algum *software* que contenha essa opção, como por exemplo, o RStudio ou Mplus. Verifica-se, então, a correlação entre os traços latentes mensurados por ambos os instrumentos e espera-se encontrar correlações significativas e de relevância em magnitude entre atributos semelhantes avaliados pelas duas medidas. É possível realizar esta análise utilizando um instrumento no formato Likert, no entanto, é importante destacar é que provável que as correlações não sejam tão altas devido à diferença no método de formato de resposta.

### Considerações finais

O presente artigo teve como objetivo fomentar a discussão sobre a construção de instrumentos de escolha forçada do tipo RANK, trazendo informações sobre a forma com a qual é feita a organização dos itens nos blocos de respostas, apresentando exemplos, sugestões de *software* e as equações matemáticas que embasam tais cálculos. Há uma carência de artigos sobre medidas ipsativas no Brasil. Mais ainda, há carência de instrumentos brasileiros nesse formato. Assim, este trabalho visa contribuir para diminuir estas lacunas.

Sugere-se que seja adotada a Teoria de Resposta ao Item Thurstoniana (TRI-T) com base nos autores Brown e Maydeu-Olivares (2011, 2012, 2013, 2018) e Brown (2014). Ainda, sugere-se que a condução da criação dos blocos de itens deste tipo de instrumento seja feita em seis grandes etapas, (a) definição do construto, (b) escrita dos itens da escala, (c) busca de evidências de

validade de conteúdo, (d) organização dos itens para compor os blocos de escolha forçada, (e) busca de evidências de validade baseadas na estrutura interna e, por fim, a (f) busca de evidências de validade baseadas na relação com outras variáveis, chamando atenção para as ações que envolvem a análise da desejabilidade social dos juízes leigos, a escrita dos itens no formato positivo e negativo, a composição dos blocos com itens com médias de desejabilidade social semelhante, dentre outros pontos apresentados sobre as análises dos dados.

É importante salientar que não foi possível contemplar todos os aspectos referentes à construção de um instrumento de escolha forçada, tampouco foi possível discutir em exaustão os tópicos apresentados. Entretanto, espera-se que as informações aqui presentes possam auxiliar os pesquisadores no desenvolvimento de instrumentos de escolha forçada do tipo RANK, considerando o conteúdo aqui apresentado de modo a favorecer e estimular mais pesquisas contemplando instrumentos deste tipo.

## Referências

Alexandre, N. M. C., & Coluci, M. Z. O. (2011). Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. *Ciência & Saúde Coletiva*, 16(7), p. 3061-3068. <https://doi.org/10.1590/S1413-812320110000800006>

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). Standards for educational and psychological testing (6. ed.). American Educational Research Association.

Borsa, J. C., & Size, M. M. (2017). Construção e Adaptação de Instrumentos Psicológicos: dois caminhos possíveis. In B. F. Damásio, & J. C. Borsa, *Manual de Desenvolvimento de Instrumentos Psicológicos* (pp. 15-37). Vetor.

Brown, A. (2014). Item Response Models for Forced-Choice Questionnaires: A Common Framework. *Psychometrika*, 81(1), 135-160. <https://doi.org/10.1007/s11336-014-9434-9>

Brown, A., & Maydeu-Olivares, A. (2011). Item Response Modeling of Forced-Choice Questionnaires. *Educational and Psychological Measurement*, 71(3), 460-502. <https://doi.org/10.1177/0013164410375112>

Brown, A., & Maydeu-Olivares, A. (2012). Fitting a Thurstonian IRT model to forced-choice data using Mplus. *Behavior Research Methods*, 44(4), 1135-1147. <https://doi.org/10.3758/s13428-012-0217-x>

Brown, A., & Maydeu-Olivares, A. (2013). How IRT Can Solve Problems of Ipsative Data in Forced-Choice Questionnaires. *Psychological Methods*, 18(1), 36-52. <https://doi.org/10.1037/a0030641>

Brown, A., & Maydeu-Olivares, A. (2018). Modeling forced-choice response formats. In P. Irwing, T. Booth, & D. Hughes (Eds.), *The Wiley Handbook of Psychometric Testing* (pp. 1-64). John Wiley & Sons.

Cao, M., & Drasgow, F. (2019). Does forcing reduce faking? A meta-analytic review of forced-choice personality measures in high-stakes situations. *Journal of Applied Psychology*, 104(11), 1347-1368. <https://doi.org/10.1037/apl0000414>

Carvalho, L. F., & Ambiel, R. A. (2017). Construção de Instrumentos Psicológicos. In B. F. Damásio, & J. C. Borsa, *Manual de Desenvolvimento de Instrumentos Psicológicos* (pp. 39-55). Vetor.

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological assessment*, 31(12), 1412-1427. <https://doi.org/10.1037/pas0000626>

Damásio, B. F., & Dutra, D. F. (2017). Análise Fatorial Exploratória: um tutorial com o software factor. In B. F. Damásio, & J. C. Borsa, *Manual de Desenvolvimento de Instrumentos Psicológicos* (pp. 241-265). Vetor.

Franco, V. R., Valentini, F., & Iglesias, F. (2017). Introdução à Análise Fatorial Confirmatória. In B. F. Damásio, & J. C. Borsa, *Manual de Desenvolvimento de Instrumentos Psicológicos* (pp. 295-322). Vetor.

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. <https://doi.org/10.1348/000711006X126600>

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. *Structural Equation Modeling*, 25(4), 621-638. <https://doi.org/10.1080/10705511.2017.1402334>

Hernández-Nieto, R. A. (2002). *Contributions to statistical analysis*. Universidade de Los Andes.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.

Pacico, J. C. (2015). Como é Feito um Teste? Produção de Itens. In Hutz, C. S., Bandeira, D. R., & Trentini, C. M. (Orgs.) *Psicometria* (pp. 55-70). Artmed.

Petrides, K. V., Furnham, A., & Mavroveli, S. (2007). Trait emotional intelligence: Moving forward in the field of EI. In G. Matthews, M. Zeidner, & R. D. Roberts (Eds.), *Series in affective science. The science of emotional intelligence: Knowns and unknowns* (pp. 151-166). Oxford University Press.

Polit, D. F., & Beck, C. T. (2006). The Content Validity Index: are you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29, 489-497.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org>

Roger-Welter, G. M. (2014) *Teste HumanGuide – Manual* (Vol. 1.).

Roger-Welter, G. M., & Capitão, C. G. (2007). Medidas Ipsativas na Avaliação Psicológica. *Avaliação Psicológica*, 6(2), 157-165. [http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1677-04712007000200006&lng=pt&tlng=pt](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712007000200006&lng=pt&tlng=pt)

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing reliability. *Psychological Bulletin*, 86(2), 420-428.

Siegel, S., & Castellan, H. J. (2006). *Estatística não paramétrica para ciências do comportamento* (2. ed.). Artmed.

Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for "optimal" cutoff properties: Fit index criteria in structural equation modeling. *The Journal of Experimental Education*, 74(3), 267-288. <https://doi.org/10.3200/JEXE.74.3.267-288>

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273-286. <https://doi.org/10.1037/h0070288>

Thurstone, L. L. (1931). Rank order as a psychophysical method. *Journal of Experimental Psychology*, 14(3), 187-201. <https://doi.org/10.1037/h0070025>

Valentini, F. (2018). Modelo latente para itens de escolha forçada. *Avaliação Psicológica*, 17(1), 1-2. <http://dx.doi.org/10.15689/ap.2017.1701.ed>

Zacharias, J. J. M. (2000). *QUATI: Questionário de Avaliação Tipológica (versão II)*. Vetor.

---

## Paloma Pereira de Almeida

Doutora em Psicologia pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), no Rio de Janeiro, RJ, Brasil; mestre em Psicologia pela Universidade Federal de Pernambuco (UFPE), em Recife, PE, Brasil. Professora da Fundação Getúlio Vargas (FGV), no Rio de Janeiro, RJ, Brasil; Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), no Rio de Janeiro, RJ, Brasil; e Universidade de São Paulo / Escola Superior de Agricultura Luiz de Queiroz (USP/Esalq), em Piracicaba, SP, Brasil. Consultora da ASEleta – consultoria especializada em gestão de pessoas e líderes.

---

## Juliane Callegaro Borsa

Doutora em Psicologia pela Universidade Federal do Rio Grande do Sul (UFRGS), em Porto Alegre, RS, Brasil; mestre em Psicologia pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), em Porto Alegre, RS, Brasil. Psicóloga em consultório privado.

---

## Jesus Landeira Fernandez

Doutor em Neurociências e Comportamento pela University of California at Los Angeles (UCLA), em Los Angeles, Estados Unidos; mestre em Psicologia pela Universidade de São Paulo (USP), em São Paulo, SP, Brasil. Professor Titular do Departamento de Psicologia da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), no Rio de Janeiro, RJ, Brasil. Pesquisador do CNPq e Cientista do Nosso Estado pela FAPERJ.

*Os textos deste artigo foram revisados pela SK Revisões Acadêmicas e submetidos para validação do(s) autor(es) antes da publicação.*