



ESTUDO DO INSTRUMENTO DE VALIDAÇÃO DE ITENS DE QUESTIONÁRIO APLICADO À PSICOLOGIA DA SAÚDE

Marília de Freitas Lima ¹
Edwirde Luiz Silva Camêlo ²

RESUMO

Este trabalho apresenta um breve estudo sobre itens de um questionário aplicado em psicologia da saúde. O modelo de resposta ao item dicotômicos sobre o nível de stress de indivíduos. Apresenta um dos aspectos fundamentais para o estudo das Teorias de Resposta ao Item que a dificuldade. Mostra de forma simples um estudo da dificuldade dos itens de um questionários aplicado em psicologia da saúde. Apresenta o índice de dificuldade do item e sua fórmula bem como correção proposta. O principal instrumento de validação de um questionário usado foi a dificuldade. Estuda-se: dificuldade dos itens, índice de dificuldade do item, dificuldade e média do teste, dificuldade e variância do teste, acertos por azar e suas correções e, por último a dificuldade e seleção dos itens. Conclui-se, portanto, que na construção do teste é recomendável por maior quantidade de itens com dificuldade média e uma pequena quantidade de itens com dificuldade alta e outra com dificuldade baixa.

Palavras-chave: Instrumento, Validação, Dificuldade, Itens.

INTRODUÇÃO

O desenvolvimento da teoria da resposta ao item (TRI) atingiu um ponto em que diversas aplicações em programas de teste educacional ou psicológico (Fajrianthi, 2017) ou em pesquisa, podem ser realizadas inteiramente com métodos de TRI. No entanto, a TRI só começou a ser aplicada há pouco tempo no campo dos instrumentos de resultados de saúde (Sekely, 2018). De acordo com pesquisas anteriores, os métodos TRI têm vantagens óbvias em comparação com a teoria de teste clássica. Uma distinção crucial entre essas duas é que a TRI define uma escala para a variável potencial medida por um conjunto de itens, e os itens são calibrados como para a mesma escala. O presente trabalho apresenta uma base da Teoria de Resposta ao Item, que se destina a importante do estudo do nível de dificuldade no item aplicado a psicologia da saúde.

¹ Mestranda do Programa de Pós-Graduação em Psicologia da Saúde da Universidade Estadual da Paraíba - UEPB, marilialimapsi@gmail.com;

² Professor do Programa de Pós-Graduação em Psicologia da Saúde da Universidade Estadual da Paraíba - UEPB, edwirde@servidor.uepb.edu.br;



Pode-se dizer que a Teoria de Resposta ao Item (TRI) é um conceito básico de medição e usando ferramentas estatísticas e matemática, busca encontrar uma descrição teórica para explicar o comportamento dos dados empíricos derivados da aplicação de um instrumento psicométrico. Os parâmetros estimados pelo modelo permitem avaliar a qualidade da técnica de cada um dos itens separadamente e do instrumento como um todo e ao mesmo tempo estima o nível que cada indivíduo apresenta um tópico de interesse. Em um modelo TRI é assumido que existe uma variável latente (θ), não diretamente observável e que se pretende estimar para cada respostas examinando o instrumento de medição.

O índice de dificuldade (ID) ou parâmetro de localização do item i corresponde ao valor θ necessário ter probabilidade de $0,5+IC/2$ de responder a resposta chave do item i , sendo IC o intervalo de confiança. Como o parâmetro θ , pode variar de menos infinito a mais infinito, embora, na prática, seus valores geralmente estão dentro do intervalo $(-4,4)$. No contexto de medição de personalidade é descrita como o ponto de transição (na escala de traço) entre a probabilidade de considerar a declaração do item como não descritivo do avaliado e considerá-lo como descritivo (Richaud, 2005).

A análise dos itens é definida como o estudo dos parâmetros cujas características estão relacionadas com as propriedades e a finalidade última do teste. Melhorar os itens, ou eleger os mais adequados, apenas tem como finalidade de melhorar as propriedades psicrométricas do teste.

Segundo (Figaldo, 2005) a análise dos itens ajuda ao construtor do teste a escolher os melhores itens, de entre um conjunto deles muito maior de que seria necessário para que o teste final resulte adequado. Ou seja, a análise dos itens permite ao pesquisador decidir que itens são pertinentes e quais não são em função da finalidade e do objeto de medida do teste total. Pode-se afirmar ainda, que a análise dos itens é um processo de seleção: apenas retém os melhores itens o resto será rejeitado.

Alves et al (2017), afirmam que a compreensão da psicologia da saúde deve ser de uma disciplina autônoma, mas essencialmente interdisciplinar, visto que se desenvolve sobre uma base multi e interdisciplinar, pois envolve saberes e práticas oriundas de outras disciplinas, como: a psicologia social e comunitária, a psicologia clínica, a saúde pública, a epidemiologia, a antropologia, a sociologia, a medicina, entre outras.



Sendo assim, este artigo possui como objetivo desenvolver uma explicação para o funcionamento da validação por meio do cálculo de dificuldade e suas estatísticas e correções aplicadas a questionários em psicologia da saúde.

METODOLOGIA

Apresenta de forma clara e simples algum conhecimento essencial para construção de questionário com resposta ao item. O principal instrumento para validação do questionário é o índice de dificuldade e suas variações a média, variância e aceitos por azar e sua correção. Serão usadas as funções `irt.item.diff.rasch(items)` e `irt.discrim(item.diff,theta,items)` dos pacotes `ltm` (Rizopoulos, 2006) e `psych` (Revelle, (2019)). Essas funções estimam as dificuldades do item usados no contexto de TRI.

REFERENCIAL TEÓRICO

São muitas as características observadas nos itens, no entanto, se observará aqueles aspectos relacionados com as propriedades dos testes. Para avaliar a qualidade dos itens os indicadores mais utilizados são: dificuldade, discriminação, análise dos detratores, fiabilidade, validade e dimensionalidade.

O grau de dificuldade de um item está em função do número de elemento da pessoa que respondem de forma correta ao mesmo.

$$ID = \frac{PA}{N}$$

PA - é o número de pessoas que de depressão que acertam o item

N - é o número total de pessoas que tem depressão e que tentam responder ao item

Em itens de eleição múltipla, para corrigir os efeitos do azar utiliza-se o índice de dificuldade corrigido:

$$ID' = p - \frac{q}{c - 1} = \frac{PA - \frac{Er}{c - 1}}{N}$$

Em que: p = proporção de acertos (índice de dificuldade sem correção); q = proporção de falhas; c = número de alternativas do item; PA = acertos; Er = erro e N = total de indivíduos.

Em psicologia da saúde pode-se dizer que dificuldade é algo relativo, depende do número de pessoas que respondem ao teste, observa-se o N no denominador da fórmula ID'.

Os valores do índice de dificuldade de um item são frequentemente classificados por:

Quadro 1: Índice de dificuldade

Muito fácil	$ID \leq 0,75$
Fáceis	$0,55 \leq ID < 0,75$
Normais	$0,45 \leq ID < 0,54$
Difíceis	$0,25 \leq ID < 0,44$
Muito difíceis	$ID < 0,25$

A dificuldade e média do teste será:

$$IDm = \frac{\sum_{i=1}^n T_i}{N}$$

Em que N é o número de intentaram responderem ao item

Supondo que o valor de T (sendo T o número de itens) do teste apenas pode ser 1 ou 2 segundo se tem acertado ou errado o item. É evidente que a média das pontuações do teste seja igual a soma dos índices de dificuldade dos itens, ou seja

$$\bar{T} = \sum_{i=1}^n ID_i \quad (1)$$

Observa-se também que indireto a relação da variância com o índice de dificuldade: se todos os participantes acertam ou erra o item, seu índice de dificuldade será zero ou um, e sua variância nestes extremos seria 0, isso é claro, pois se não existe variação não existe variância. Neste caso o item resultará pouco útil como instrumento de medida em psicologia da saúde. Como consequência, qualquer sistema de classificação ou escalamento dos indivíduos que respondem ao item resultaria inviável, ou seja, para que um item seja útil é imprescindível que gere variabilidade, dispersão, nos indivíduos que o respondem (Thorndike, 1989). É importante que as respostas não sejam todas iguais aos indivíduos, apenas através das diferenças entre as respostas será possível estatisticamente classificar os indivíduos.



A relação entre a dificuldade dos itens e a variância do teste quando os itens forem dicotômicos será dada por:

$$S_x^2 = p \cdot q = p(1 - p) = p - p^2$$

Em que p é a proporção de acertos e q , proporção de não acertos, ou seja, $q = 1 - p$. O valor de $S_x^2 = p \cdot q$ é possível quando o item é dicotômico.

Para calcular o máximo de uma função é necessário calcular a primeira e a segunda derivada da dita função. Observa-se que a segunda derivada é negativa, o máximo é encontrado igualando a zero. Derivando a expressão anterior em função de p tem-se:

$$\theta'_p = 1 - 2p \quad \text{e} \quad \theta''_p = -2$$

Igualando a primeira derivada à zero, tem-se: $0 = 1 - 2p \rightarrow p = 0,5$. Ou seja, é máximo para $p = q = 1/2$, já que a derivada segunda é negativa.

Com esse resultado pode-se concluir que a variância de um item dicotômico é máxima quando seu índice de dificuldade é 0,5 (Normais, pois pertence ao intervalo $0,45 \leq ID < 0,54$). Na realidade o índice de dificuldade deveria ser chamado de índice de facilidade. Observa-se que é uma função crescente que vai desde valores baixos quando a dificuldade é maior, até valores altos quando a dificuldade é maior.

Isso significa que a variância de um teste formado por itens dicotômicos estará tanto mais próximo de seu máximo quando mais próximo esteja a 0,5 os índices de dificuldade de seus itens. Assim, as possibilidades de estabelecer as máximas discriminações entre os indivíduos que respondem ao um teste dependem diretamente da dificuldade de seus itens. Neste sentido fica justificada a afirmação de Wang e Hung quando indicam que a dificuldade é um dos aspectos mais importantes dos itens de um teste (Wang e Hung, 2002). Suponha que a distribuição da variável V tenha média 0,50 e variância 0,7 para a primeira resposta (correta). O indivíduo escolhe aquela resposta cujo valor da variância seja maior, como se observa na Figura 1. Existe um afastamento do centro (média) a medida que as respostas aos itens variam, logo a curva passa de um decrescimento cada vez mais acentuado para um decrescimento cada vez menos acentuado. Se o índice de dificuldade for próximo de 0,5 haverá uma menor dispersão em relação na resposta dos itens.

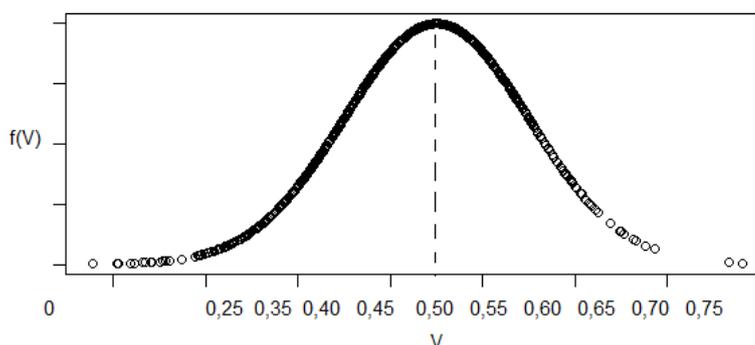


Figura 1. Distribuição dos processos discriminantes

O índice de dificuldade, em sentido estrito de indicador do difícil que pode ser acertar ou errar um item, apenas tem sentido calcular nos testes em que existam respostas corretas ou incorretas, não nos testes de personalidade ou escalas de atitudes (Muñiz J, et al, 2005).

Por outro lado, nos itens que, pela natureza do que se pergunta em psicologia da saúde, como sempre, existem repostas corretas e incorretas que depende não apenas do difícil que seja o item e do nível de conhecimento sobre o tema proposto que tenha quem resposta, senão também do número de alternativas que o item e do nível tenham. Se um questionário tem duas alternativas de respostas, independentemente do que se pergunte sobre psicologia aplicada à saúde, o índice de dificuldade deste item será 0,5, já que a metade dos indivíduos que tentam responder pensando no acerto por sorte. Para evitar isso, calcula-se o índice de dificuldade ajustado para minimizar o efeito da sorte em itens deste tipo. Essa correção é dada pela seguinte equação:

$$I.D.' = p - \frac{q}{c - 1}$$

Em que I.D.' é o índice de dificuldade ajustado; p é a proporção de acertos (índice de dificuldade sem correção); q é a proporção de não acertos e c é o número de alternativas que tem o item.

RESULTADOS E DISCUSSÃO

A seguinte tabela pode servir para servir como exemplos de itens em psicologia da saúde. Em que os 10 indivíduos responderam um questionário em que os item foram 1 =

muito pouco estresse, 2 = pouco estresse, 3 = estressado, 4 = muito estresse e 5 = elevado nível de estresse.

Tabela 1. Indivíduos, itens e o índice de dificuldade

Indivíduos	Item 1	Item 2	Item 3	Item 4	Item 5	Total
Ind. 1	1	0	1	1	0	3
Ind. 2	0	0	1	0	0	1
Ind. 3	1	0	0	0	0	1
Ind. 4	1	0	0	0	0	1
Ind. 5	1	0	0	0	1	2
Ind. 6	1	1	0	0	1	3
Ind. 7	1	1	1	0	1	4
Ind. 8	1	1	0	1	1	4
Ind. 9	0	1	0	0	1	2
Ind. 10	0	1	0	0	1	2
Índice de dificuldade (ID)	0,7	0,5	0,3	0,2	0,6	SP=2,3

Em que SP é igual à soma da proporção de acerto e também a média da quantidade de acertos nos itens.

De acordo com a equação (1) tem-se: $\bar{T} = \frac{3+1+1+1+2+3+4+4+2+2}{10} = 2,3$, que é igual a soma dos índice de dificuldade $\sum_{i=1}^n ID_i = 0,7 + 0,5 + 0,3 + 0,2 + 0,6 = 2,3$. Apenas o item um teve um índice de dificuldade muito fácil, já o item 4 apresenta um índice de dificuldade difícil.

Considere agora um caso em que se têm apenas três itens para exemplificar o cálculo do índice de dificuldade. Como se observa a diferenças são apreciáveis.

Tabela 2. Indivíduos, itens e os índices de dificuldades

Indivíduos	Item 1	Item 2	Item 3
Ind1	1	0	1
Ind2	0	0	1
Ind3	1	0	0
Ind4	1	0	0
Ind5	1	0	0
Ind6	1	1	0
Ind7	1	1	1
Ind8	1	1	0
Ind9	0	1	0
Ind10	0	1	0
I.D.	0,7	0,5	0,3
I.D*	0.5	0,25	0,05

Observa-se que no item 3 uma acentuada diferença no índice de dificuldade ($0,3 - 0,05 = 2,5$), já no item 1 a diferença foi de 0,2. No item 1 o item é considerado fácil; no item 2, os dois índices de dificuldades variam entre difícil e fácil; e no item 3, varia de difícil e muito difícil. Concordando com as afirmações de Sartes e Formigoni (2013) em que testes distintos com índices de dificuldade e discriminação diferentes produzem resultados diferentes para os mesmos indivíduos.

CONSIDERAÇÕES FINAIS

Nesse trabalho se propôs, como objetivo geral, elaborar de forma explicação o funcionamento da validação através do cálculo de dificuldade e suas estatísticas e correções aplicadas a questionários em psicologia da saúde. A compreensão dos itens deve expresso de modo que facilite aos indivíduos o entendimento das perguntas. Na construção do teste é aconselhável por maior quantidade de itens com dificuldade media (aqueles com histograma simétrico) e uma pequena quantidade de itens com dificuldade alta e outra com dificuldade



baixa. Entendendo que o índice de dificuldade (em realidade deveria se chamado índice de facilidade) que é uma função crescente que vai desde valores baixos, quando apresenta maior dificuldade, até valores altos, como por exemplo, 0,9 quando respondem corretamente a 90% dos indivíduos. Para construção de questionário é aconselhável colocar maior quantidade de itens com dificuldade média (aqueles com histograma simétrico) com índice de dificuldade 0,5 e uma pequena quantidade de itens com dificuldade alta e outra com dificuldade baixa.

REFERÊNCIAS

- ALVES, B.C.A.; ANGELONI, R.V.; AZZALIS, L.A.; PEREIRA, E.C.; PERAZZO, F.F.; ROSA, P.C.P.; FEDER, D.; JUNQUEIRA, V.B.C.; FONSECA, F.L.A. Esclerose múltipla: revisão dos principais tratamentos da doença. *Saúde Meio Ambiente*, 3 (2): 19-34, 2014.
- CAMPBELL, N.R. **Na account of the principles of measurement and calculations**. London: Longmans Green.
- CAMPBELL, N.R. **Symposium: Measurement and its importance for philosophy**. *Proceedings Aristot. Society Suppl.*, 17,121-142. London: Harrison.
- FAJRIANTHI, Z. R. A. Development of a psychological test to measure ability-based emotional intelligence in the Indonesian workplace using an item response theory. *Psychology Research & Behavior Management* 10, 339–352 (2017).
- MIYAZAKI, M.C.O.S., DOMINGOS, N.A.M., CABALLO, V.E. Psicologia da Saúde: intervenções em hospitais públicos. In: B. Rangé (org.). **Psicoterapias Cognitivo-Comportamentais: um diálogo com a psiquiatria**, (pp.463-474). Porto Alegre: Artmed, 2001.
- MUÑIZ,J., ÁNGEL M. FIDALGO, EDUARDO GARCÍA-CUETO, RAFAEL, MARTINEZ E RAFAEL, MORENO. **Análise de los ítems, Cuadernos de Estadística**. Madrid: Editorial La Muralla, S.A., 2005.
- PASQUALI, L. **Psicometria – teoria dos testes na Psicologia e na Educação**. Petrópoles, RJ: Vozes, 2003.
- REVELLE, W. (2019) **psych: Procedures for Personality and Psychological Research**, Northwestern University, Evanston, Illinois, USA. Disponível em: <https://CRAN.R-project.org/package=psych> Version = 1.9.12. Acesso em: 17 nov 2020.
- RIZOPOULOS, D. (2006). Itm: **An R package for Latent Variable Modelling and Item Response Theory Analyses**, *Journal of Statistical Software*, 17 (5), 1-25. Disponível em: <http://www.jstatsoft.org/v17/i05/> . Acesso em: 17 nov 2020.
- RICHAUD, M. C. (2005). **Desarrollos del análisis factorial para el estudio de ítems dicotómicos y ordinales**. *Interdisciplinaria*, 22, 237 – 251.
- SARTES, Laisa Marcorela Andreoli; SOUZA-FORMIGONI, Maria Lucia Oliveira de. Avanços na psicometria: da Teoria Clássica dos Testes à Teoria de Resposta ao Item. **Psicol. Reflex. Crit.**, Porto Alegre , v. 26, n. 2, p. 241-250, 2013 . Available from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-79722013000200004&lng=en&nrm=iso Acesso em: 17 Nov. 2020. <https://doi.org/10.1590/S0102-79722013000200004>.



SEKELY, A., Taylor G. J. & Bagby, R. M. Developing a short version of the Toronto Structured Interview for Alexithymia using item response theory. **Psychiatry Research**, 2018.

THORNDIKE, R.L. **Psicometria aplicada**, Maxico, 1989.

WANG, W.C. E HUNG, L.F. **Na objective procedure for determining IF items have good discrimination**. Chinese Journal of Psychology, \$\$ (2), p. 253-262.