

TRABALHO FINAL DE TEORIAS DE AVALIAÇÃO

Amábile Galdino Leandro - GRR20185612

Gabriel José Cavassin Fabri - GRR20195780

DESCRIÇÃO E OBJETIVO

“A necessidade de uma teoria de testes e medidas corre como consequência do reconhecimento de que as provas são por natureza falhas, porém, passíveis de aperfeiçoamento” (FLETCHER, 2010, p. 7).

Fletcher traz que essas teorias possibilitam uma melhor interpretação sobre o conhecimento dos respondentes. Nesse sentido, este trabalho possui o objetivo de apresentar uma análise de um conjunto de respostas relacionadas ao conhecimento sobre filmes (traço latente) a partir de duas metodologias de avaliação, a Teoria Clássica dos Testes (TCT) e o modelo de ajuste de 3 parâmetro da Teoria de Resposta ao Item (TRI).

DESCRIÇÃO DAS METODOLOGIAS

O conjunto de dados foi resultado de um trabalho coletivo, o qual os estudantes da disciplina de Teorias de Avaliação elaboraram os itens a partir de 10 descritores para gêneros de filme: (D1) ação/aventura; (D2) animação; (D3) comédia; (D4) documentário/biografia; (D5) drama; (D6) ficção científica; (D7) musical/dança; (D8) policial/guerra, (D9) seriados; e (D10) terror/suspense, produzindo 44 itens, buscando seguir as diretrizes de elaboração de itens do INEP (BRASÍLIA, 2010). Assim, foi realizada a avaliação desses e após o processo, restaram 13 itens para a composição do instrumento de medida, que foi submetido a 1148 respondentes, mantendo apenas 1050 depois do tratamento da base de dados.

Para fundamentar a Teoria Clássica dos Testes foram utilizados os autores Fletcher (2010) e Borgatto e Andrade (2012), buscando apresentar os preceitos teóricos, a organização dos métodos da TCT e os resultados das análises. Enquanto o ajuste do modelo de três parâmetros foi orientado pelos trabalhos Andrade, Tavares e Valle (2000) e Fletcher (2010), para desenvolver a calibração dos itens e analisar as Curvas Características dos Itens. E quanto a construção da escala foram utilizados os autores Andrade, Tavares e Valle (2000) e Barbeta (2014).

As duas metodologias foram realizadas com o auxílio dos *softwares* Microsoft Excel (2016) e R, versão 4.0.2 (2020) utilizando os pacotes “lrm”, “CTT”, “mirt”, “readxl”, “irtos”, em busca de obter os resultados para o desenvolvimento das análises da TCT, do ajuste de 3 parâmetros e da construção da escala dos itens.

ANÁLISES A PARTIR DA TEORIA CLÁSSICA DOS TESTES

De acordo com Fletcher (2010), a TCT fornece um método de descrição para a variação esperada da pontuação de cada indivíduo avaliado sobre sua verdadeira pontuação, logo se trata de uma análise pautada apenas nas pontuações. Borgatto e Andrade (2012) colocam ainda que esses parâmetros descritivos auxiliam a interpretação da distribuição das respostas nos itens. Dessa forma, atenta-se a dois aspectos não observáveis na TCT, o erro e o escore verdadeiro. Nesse sentido, é relevante para TCT observar as frequências dos escores, de modo a estabelecer um panorama geral de distribuição de acertos sobre o traço latente.

Tabela 1: Frequências de escore

Escore	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Frequência	1	2	3	7	14	45	73	95	149	160	194	144	123	40

Fonte: Autores (2020)

Em geral causa a impressão de que se trata de um teste fácil, pois as maiores frequências estão concentradas nos escores superiores. Nesse sentido, o processo da análise clássica, para além do número de acertos, também é necessário analisar o questionário como um todo, sendo necessário definir alguns parâmetros que constituem a **Tabela 2**.

Tabela 2: Parâmetros da Teoria Clássica de Teste

Itens	Erros	Acertos	Coef. ponto-bisserial	Alfa de Cronbach
1	0,2838	0,7162	0,4322	0,5921
2	0,3571	0,6429	0,3745	0,6090
3	0,4971	0,5029	0,5342	0,5707
4	0,0952	0,9048	0,4003	0,5896
5	0,3524	0,6476	0,5480	0,5653
6	0,8190	0,1810	0,2858	0,6139
7	0,2124	0,7876	0,3042	0,6136
8	0,3343	0,6657	0,5218	0,5719
9	0,1971	0,8029	0,5491	0,5629
10	0,3952	0,6048	0,4010	0,6042
11	0,1543	0,8457	0,4148	0,5890
12	0,1495	0,8505	0,3865	0,5937
13	0,0429	0,9571	0,2848	0,6026

Fonte: Autores (2020)

Nessa tabela são apresentadas as proporções de erros e acertos dos respondentes, e outros dois parâmetros que caracterizam os itens e o instrumento de avaliação. O primeiro deles é o parâmetro de discriminação, também conhecido como o coeficiente ponto-bisserial.

Analizando o coeficiente ponto-bisserial, os itens 6 e 13 estão abaixo dos valores considerados adequados, apresentam valores menores que 0.3, logo é indicado que esses itens sejam retirados no processo de avaliação.

O parâmetro alfa de Cronbach diz respeito à consistência interna do instrumento de medida, que quanto mais próximo de 1, mais adequado (BORGATTO e ANDRADE, 2012). Além dos alfas de cada item, tem-se também o mesmo parâmetro para o questionário como um todo, que nesse caso seria de 0.611, o que caracteriza que o instrumento de medida não é ideal (aqueles que possuem alfa de Cronbach superiores a 0.7).

Sabe-se que “o modelo clássico parte do pressuposto de que o escore de uma prova é por natureza falha e contém alguma parcela de erro. Logo, se uma parte do resultado é de erro, o restante deve ter sua base real ou verdadeiro”. (FLETCHER, 2010, p. 13). Uma outra forma de avaliar seria utilizando a Teoria de Resposta ao Item.

ANÁLISES A PARTIR DO AJUSTE DO MODELO DE 3 PARÂMETROS

O modelo de 3 parâmetros é um dos propostos dentro da Teoria de Resposta ao Item (TRI), essa “[...] baseia-se em modelos que representam a probabilidade de um aluno responder corretamente a um item como função dos parâmetros do item e da(s) habilidade(s) do respondente.” (ANDRADE e VALLE, 1998, p. 14). Assim, é possível formar uma escala de habilidade dos respondentes (θ) que pode representar a proficiência do indivíduo. Esses autores colocam que o modelo logístico unidimensional de 3 parâmetros é ideal para itens de múltipla escolha dicotomizados, aqueles que podem ser caracterizados entre certo ou errado, como os itens em estudo. Conforme Andrade e Valle (1998), esses três parâmetros consistem em: (a) - coeficiente de discriminação ou inclinação proporcional da Curva Característica do Item (CCI); (b) - parâmetro de dificuldade do item que possui a mesma escala que a habilidade; (c) - parâmetro de acerto casual do item.

Primeiramente é realizado o processo de calibração dos itens, de forma a estabelecer se esses itens são adequados para testar as habilidades dos respondentes. Como visto na seção sobre a Teoria Clássica, os itens 6 e 13 foram retirados na análise do coeficiente ponto-bisserial. No processo de calibração, apenas o item 7 apresentou discrepância no parâmetro de discriminação ($a = 0.398 < 0.6$), sendo retirado em seguida. Dessa forma, restaram os itens:

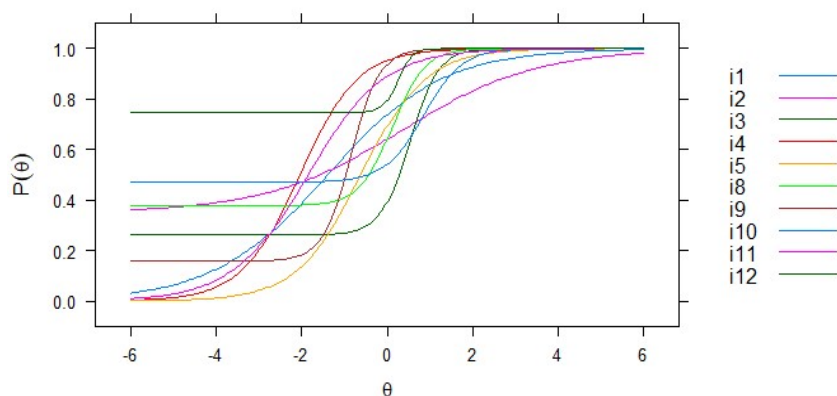
Tabela 03: Itens calibrados

Itens	a	b	c
1	0,743	-1,392	0,000
2	0,628	0,343	0,350
3	2,960	0,524	0,261
4	1,459	-2,048	0,001
5	1,316	-0,611	0,000
8	2,635	0,118	0,378
9	2,982	-0,839	0,158
10	2,178	0,849	0,471
11	1,122	-1,855	0,000
12	6,038	0,248	0,747

Fonte: Autores (2020)

Dentre os itens restantes, o item 12 se destaca pelo fato de apresentar um parâmetro a discrepante em relação aos demais itens, caracterizando-o como um item que não distingue as habilidades dos respondentes, porém foi mantido respeitando as condições impostas pelo professor para o desenvolvimento do trabalho. Para reconhecer um panorama geral, a **Figura 1** representa as Curvas Características do Itens calibrados, lembrando que “a curva de característica do item é uma função matemática que relaciona a probabilidade de um acerto na resposta do item a um valor específico de θ encontrado em algum ponto ao longo da escala da habilidade latente” (FLETCHER, 2010, P. 26)

Figura 1: Curvas Características dos Itens calibrados



Fonte: Autores (2020)

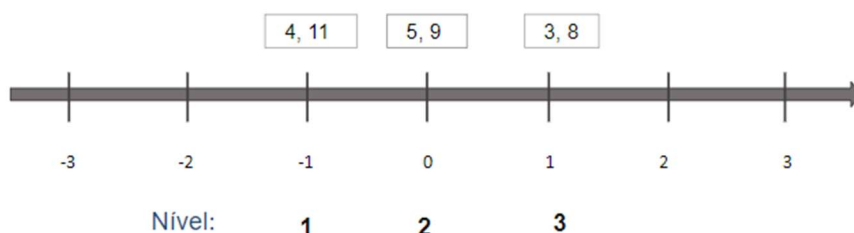
Com o auxílio visual das CCI pode-se confirmar a inadequação do item 12, pois se trata de um item que não discrimina bem as habilidades dos respondentes. Agora, com base nas funções de probabilidade dos itens fornecidas pelo método, na próxima seção é apresentado a construção e interpretação de uma escala.

CONSTRUÇÃO E INTERPRETAÇÃO DE ESCALA

Para fazer o posicionamento dos itens, a escala foi construída baseada no trabalho de Barbetta (2014), a partir de uma distribuição normal com média zero e variância unitária, com os itens de maior discriminação posicionados em um intervalo de -3 a 3.

Andrade, Tavares e Alves (p. 110, 2000) definem como itens âncora para níveis consecutivos Y e Z, com $Y < Z$, se e só se 3 condições forem satisfeitas simultaneamente: **(1)** $P(U = 1|\theta = Z) \geq 0.65$; **(2)** $P(U = 1|\theta = Y) < 0.50$; e **(3)** $P(U = 1|\theta = Z) - P(U = 1|\theta = Y) \geq 0.30$. E para os quase-âncoras, era necessário atender dois dos três critérios citados. De acordo com esses critérios, os itens 3, 5 e 9 foram considerados como âncoras, enquanto os itens 4, 8 e 11 quase-âncoras, e ao posicioná-los na escala, obtêm-se:

Figura 2: Posicionamentos dos itens em escala



Fonte: Autores (2020)

O **nível 1** caracteriza os respondentes com conhecimentos superficiais sobre produções cinematográficas, de forma que para as características avaliadas não era necessário assistir os filmes em questão. O **nível 2** evidencia um grupo de pessoas que reconheceram características particulares de filmes ou séries, como algumas partes específicas, sendo necessário conhecer as histórias presentes nas produções. O **nível 3** marca um interesse para além de apenas assistir os filmes/séries, sendo necessário interpretá-los e buscar informações específicas sobre produções cinematográficas.

Os itens reconhecidos como âncoras ou quase-âncoras demarcaram bem a distinção dos conhecimentos relacionados. Portanto, esses níveis estão condizentes com o objetivo do instrumento de medida em traçar o conhecimento sobre filmes.

CONSIDERAÇÕES FINAIS

Tanto a TCT quanto a TRI de 3 parâmetros evidenciam que o instrumento de medida é inconsistente, um por apresentar o Alfa de Cronbach menor que 0,7 e outro pelo delineamentos das CCI's. Enquanto a construção da escala, essa traduz os níveis de dificuldades sobre o conhecimento do traço latente em questão.

Em relação aos itens vale destacar que a análise consistiu em avaliar apenas o conhecimento de filmes, mas o item 7, por exemplo, apresentava no texto-base palavras estrangeiras sem traduções, o que demanda conhecimentos para além dos filmes, podendo ser a razão pela eliminação deste item. Enquanto o item 12 foi mantido apenas pelos limitantes informados para a execução do trabalho, pois a partir do ajuste de 3 parâmetros, fica evidente a falta de capacidade deste item em discriminar as habilidades dos respondentes.

Por fim, o objetivo de apresentar as interpretações dos métodos de avaliação sobre o traço latente em questão foi alcançado, pois foi possível notar as nuances de cada um dos métodos e como estes se diferem, assim como as possíveis conclusões que se pode obter a partir de cada um.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. **Teoria da Resposta ao Item: conceitos e aplicações**. 2000.

ANDRADE, D. F.; VALLE, R. C. Introdução à Teoria da Resposta ao Item: conceitos e aplicações. **Estudos em Avaliação Educacional**, n. 18, p. 13-32, 1998. Fundação Carlos Chagas. <http://dx.doi.org/10.18222/eae01819982250>

BARBETTA, P. A.; TREVISAN, L. M. V.; TAVARES, H.; AZEVEDO, T. C. A. M. Aplicação da teoria da resposta ao item uni e multidimensional. **Estudos em Avaliação Educacional**, v. 25, n. 57, p. 280-302, 2014. Fundação Carlos Chagas. <http://dx.doi.org/10.18222/eae255720142832>

BORGATTO, A. F.; ANDRADE, D. F. de. Análise clássica de testes com diferentes graus de dificuldade. **Estudos em Avaliação Educacional**, v. 23, n. 52, p. 146-156, 2012. Fundação Carlos Chagas. <http://dx.doi.org/10.18222/eae235220121934>.

BRASIL. **Guia de Elaboração e Revisão de Itens**. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. 2010.

FLETCHER, P. R. **Da teoria clássica dos teste para os modelos de resposta ao item**. Rio de Janeiro: Escola Nacional de Ciências Estatísticas, 2010.

R Core Team (Version 4.0.2). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2020. URL <https://www.R-project.org/>.