



ação ergonômica, volume 5, número 3

AValiação de Usabilidade em Sites de E-commerce: Uma Aplicação da Teoria da Resposta ao Item (TRI)

Rafael Tezza

Universidade Federal de Santa Catarina - UFSC
rafaeltezza@deps.ufsc.br

Fernando de Jesus Moreira Junior

Universidade Federal de Santa Catarina - UFSC
fmjunior@smail.ufsc.br

Antonio Cezar Bornia

Universidade Federal de Santa Catarina - UFSC
cezar@inf.ufsc.br

Resumo: O presente artigo discute os métodos convencionais de avaliação de usabilidade e apresenta uma proposta de avaliação em sites de e-commerce utilizando a Teoria da Resposta ao Item. Foi construído um questionário e aplicado em 361 sites, resultando numa escala padronizada para mensuração do grau de usabilidade em sites e e-commerce.

Palavra Chave: usabilidade, e-commerce, Teoria da Resposta ao Item.

Abstract: *This article discusses the conventional methods for evaluating usability and presents a proposal for evaluation in e-commerce sites using the Item Response Theory. We constructed a questionnaire and implemented in 361 sites, resulting in a standardized scale to measure the degree of usability in e-commerce sites.*

Keywords: *usability, e-commerce, Item Response Theory.*



1. Introdução

A literatura sobre avaliação de desempenho em comércio eletrônico (e-commerce) business to consumer (B2C) tem destacado a interação do usuário com o website como um ponto chave na decisão de compra e na fidelização do consumidor (JANDA ET AL, 2002; LONG; McMELLON, 2004; LIM; DUBINSKY, 2004; KIM; STOEL, 2004). Esta interação é tida como o primeiro contato do usuário com a empresa de e-commerce, e a qualidade de sua apresentação tem grande impacto no sucesso da mesma.

Um elemento significativo de avaliação de desempenho na economia digital é a usabilidade dos sites de e-commerce (MORANDINI, 2003). A não atenção a critérios de usabilidade em qualquer site comercial ou uma medição mal conduzida pode significar perda de clientes e/ou alocação indevida de recursos (NIELSEN; LORANGER, 2006).

Atualmente as medidas de usabilidade em sites de e-commerce vêm sendo realizadas de diversas maneiras, utilizando metodologias ou ferramentas que envolvem ou não o usuário (NIELSEN; MARCK, 1994; CYBIS, 2007). Estas medidas envolvem características objetivas e subjetivas, baseadas em critérios recomendados por especialistas ou em opiniões de usuários, gerando muitas das vezes falta de sistematização e de precisão nos resultados (CYBIS, 2007).

Para gerar um melhor entendimento das estruturas envolvidas em uma avaliação de usabilidade e para sistematizar os resultados, pode-se fazer uso de escalas de medidas alicerçadas em conceitos matemáticos e de usabilidade. Deste ponto de vista, a Teoria da Resposta ao Item (TRI) representa uma poderosa ferramenta, uma vez que esta possibilita a criação de escalas a partir de um construto que necessariamente faz uso de conceitos aprofundados

Técnica	Descrição	Vantagens	Desvantagens
Avaliação Heurística	Análise de conformidade do sistema diante de padrões de qualidade ou heurísticas definidas por especialista.	- Rapidez de aplicação, - Baixo custo, - Abrangência e - Facilidade de identificação dos problemas	- Requer conhecimento e experiências do avaliador. - São necessários de três a cinco avaliadores para um bom resultado. - Subjetiva e pouco sistemática.
Avaliação Analítica	Decomposição da tarefa para verificar a complexidade ou estimar os tempos das interações propostas	É bastante útil para comparar alternativas de interface	Não leva em consideração erros ou incidentes de interação
Lista de Verificação (Checklist)	São inspeções a requisitos propostos por padrões de qualidade baseadas em listas de verificação, <i>guidelines</i> e normas.	- Avaliação barata, rápida, sistemática e de fácil aplicação. - Outra pessoa não especializada em usabilidade pode aplicar.	- Limitada a problemas intermediários e menores. - Depende da qualidade da ferramenta (checklist).
Inspeções cognitivas	Confronto entre as lógicas de operação do projetista e de um usuário novato. Modo formalizado de imaginar os pensamentos e as ações dos usuários leigos.	- Pode ser aplicado já no início do ciclo de desenvolvimento. - Permite que analistas, designers e atuem como avaliadores. - Reconhece a forma com que o usuário executa tarefas e realiza a análise	- Não leva à identificação imediata do problema. - Requer muito tempo.
Análise da Atividade	Levantamento sobre a atividade desempenhada pelo usuário, observando-o em seu ambiente natural	- Os problemas são vistos do ponto de vista real - Identifica problemas rotineiros de uso	- Requer tempo e pessoas - Pode não ser totalmente imparcial.
Grupo Focal	Discussão entre seis a nove usuários, orientadas por um moderador.	- Promove a participação do usuário. - Boa fonte de novas idéias e/ou sugestões	- Requer tempo e pessoas - Subjetividade
Avaliação Cooperativa	Usuários e projetistas discutem juntos a interface.	- Fornece idéias para o projeto através de sugestões dos usuários. - Promove a participação do usuário. - Experiência educacional rica.	- Requer tempo e pessoas - Subjetividade
Teste de usabilidade	Simulação de situações de uso objetivando constatar problemas medir seus impactos e identificar suas causas.	- Abrangência - Verificação de situações reais de uso	- Alto custo - Requer tempo e pessoas - Pode não ser totalmente imparcial.

Quadro 1: Descrição, vantagens e desvantagens de algumas técnicas de avaliação de ergonomia de interface e usabilidade.



de usabilidade.

O objetivo deste artigo é apresentar uma nova abordagem para avaliação de usabilidade em sites de e-commerce – a Teoria da Resposta ao Item. Para isso é apresentada uma explanação geral sobre medidas de usabilidade, e sobre a TRI e por fim apresentada uma aplicação da TRI na confecção de uma escala de medida do grau de usabilidade em sites de e-commerce.

2. Avaliação de Usabilidade

O conceito de usabilidade considera a eficácia, a eficiência e a satisfação com que determinados usuários conseguem atingir objetivos específicos em determinadas circunstâncias (ISO 9241-11, 1998). Sendo assim, verifica-se a existência de três características fins, que podem determinar a avaliação de um sistema ou de uma interação, diretamente relacionada: a tarefa a ser realizada, ao ambiente, ao equipamento e ao usuário.

Desta forma, a maioria das técnicas de avaliação de usabilidade existente na literatura leva necessariamente em consideração estas características e suas relações. Adotando, evidentemente, os mais diversos pontos de vista e em alguns casos evidenciando alguma destas características e/ou fixando outras.

Os diversos métodos de avaliação de usabilidade existentes podem ser divididos entre os que são realizados sem a participação de usuários e com a participação de usuários. O quadro 1 resume algumas das técnicas utilizadas para avaliação de ergonomia de interface e usabilidade, com suas vantagens e desvantagens.

Observa-se que a maioria das técnicas possui como desvantagem a subjetividade e a necessidade de vários recursos. Por outro lado, a Teoria da Resposta ao Item possui a vantagem de reduzir a subjetividade das medidas uma vez que cria uma escala padronizada. Esta escala permite identificar tanto o desempenho da interface, ou site analisado como as características que agregam mais usabilidade e as que são mais básicas.

3. Teoria da Resposta ao Item

A Teoria da Resposta ao Item (TRI) é um conjunto de modelos matemáticos e estatísticos que procura

medir habilidades, atitudes, interesses, desempenho, maturidade, conhecimentos ou proficiência de indivíduos ou sistemas por meio de um questionário e da construção de uma escala na qual, o traço latente do respondente e a dificuldade de um item podem ser comparados (HAMBLETON, 2000). A escolha do modelo mais adequado depende, sobretudo do que se deseja medir e da natureza dos dados disponíveis.

Um dos modelos matemáticos da TRI é o modelo logístico de 2 parâmetros (ML2P), que é comumente utilizado para modelar construtos cumulativos com variáveis dicotômicas. Este modelo pressupõe a relação monótona entre o valor da variável latente (θ) e a sua probabilidade de escolha entre duas alternativas segundo uma função de distribuição logística parametrizada por coeficientes que representam as características do item:

$$P_i(\theta_j) = P(U_{ij} = 1 / \theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}$$

Onde, b_i representa a dificuldade do item i em uma determinada escala e representa o valor da variável latente θ , para o qual há 0,5 de probabilidade do indivíduo j escolher a resposta representada por $U = 1$. O coeficiente a_i é o parâmetro de discriminação do item i , proporcional à inclinação da Curva Característica do Item – CCI no ponto b_i .

A TRI cria uma escala, relacionada ao traço latente, com média zero e desvio padrão um, onde são estimados os parâmetros dos itens. Durante estas estimações, são geradas as curvas características dos itens (CCI), nas quais pode-se visualizar a probabilidade de acerto do item ($P_i(\theta_j)$) em função da habilidade do respondente (θ). A figura 1 exemplifica CCI de dois itens hipotéticos.

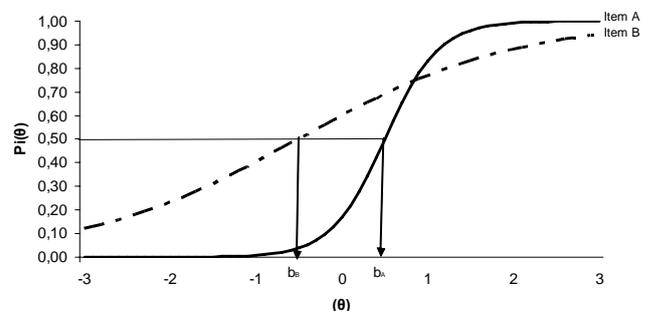


Figura 1: Curva Característica do Item (CCI)

Observa-se que o eixo y representa os valores da função de probabilidade ($P_i(\theta)$) que varia de zero a



um. No eixo x, tem-se a escala de habilidade (θ) gerada pela TRI. Nesta escala, são posicionados os itens e os respondentes. Assim, é possível comparar os desempenhos dos respondentes e a qualidade dos itens verticalmente e longitudinalmente. Pode-se visualizar, na figura 1, a representação gráfica dos parâmetros. Nota-se que o parâmetro de dificuldade (b) de um item representa a posição na escala onde a probabilidade de acerto é de 0,5. Sendo assim, o item A tem um $b = 0,5$ e o item B tem um $b = -0,5$ na escala gerada. Portanto, o item A possui um grau de exigência maior que o item B, ou seja, para um indivíduo com habilidade (θ) igual a $-0,5$ tem 50% de probabilidade de acertar o item B e 4% de probabilidade de acertar o item A. O parâmetro de discriminação (a) de um item, segundo Andrade *et al.* (2000), é proporcional à derivada da tangente da curva no ponto de inflexão, ou seja no ponto b. Verifica-se que o item A em um inclinação maior que o item B, portanto um maior poder de discriminação.

década de 1950 com Lord (1952). Entretanto, um ponto crítico na utilização da TRI é a estimação dos parâmetros, que necessita de um grande esforço computacional (ANDRADE ET AL., 2000), o que dificultou a popularização da TRI naquela época, sendo retomada recentemente em paralelo aos avanços na área computacional. No Brasil a TRI foi aplicada pela primeira vez em 1995 na área educacional na análise dos resultados do Sistema Nacional de Ensino Básico (SAEB) e no Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP) (ANDRADE ET AL., 2000).

4. Aplicação da TRI na IHC

Para a elaboração de um instrumento utilizando a TRI, é necessário primeiramente definir qual é o traço latente de interesse, ou seja, é necessário ter um conhecimento prévio do que se deseja medir. A TRI faz a análise dos itens e, a partir daí, constrói a escala padronizada. A elaboração dos itens deve contar com um conhecimento *a priori* do assunto e seguir técnicas de elaboração de itens (PASQUALI,

Os primeiros estudos sobre a TRI iniciaram-se na

Item	Descrição	Parâmetro		Item	Descrição	Parâmetro	
		a	b			a	b
1	Homepage deixa claro o que o site faz (demonstra os principais produtos e/ou uma breve descrição de seu objetivo e/ou benefícios que oferece), sem precisar usar a rolagem?	0,756	-1,55	17	É possível ampliar as fotos dos produtos para visualizar detalhes?	0,931	-0,98
2	As palavras clicáveis (coloridas ou sublinhadas) apresentam uma forma diferenciada quando são selecionadas?	0,979	-1,35	18	Em produtos em que existam mais de uma perspectiva, é possível visualizar todas as perspectivas?	0,972	1,81
3	As opções principais do site estão visíveis? Ou seja não há necessidade de passar o mouse sobre uma área gráfica para ver opções ocultas.	0,812	-2,97	19	Os grupos de botões de comando estão dispostos em coluna e à direita, ou em linha e abaixo dos objetos aos quais estão associados?	0,909	-2,90
4	A disposição dos objetos de interação de uma caixa de diálogo segue uma ordem lógica?	0,854	-0,85	20	Quando há rolagem, não existe elementos de design (na tela inicial) que pareçam com marcadores de final de página?	0,938	-3,24
5	Os rótulos de campos começam com uma letra maiúscula, e as letras restantes são minúsculas?	0,869	-1,46	21	Todos os campos e mostradores de dados possuem rótulos identificativos?	0,906	-2,40
6	O site possui opção de acesso com outras línguas?	0,79	4,65	22	O botão de finalização de compra está no final da lista?	0,752	-4,49
7	Os títulos estão alinhados a esquerda?	0,972	-2,42	23	É possível saber os custos totais antes de fazer cadastro?	0,792	-2,21
8	Os parágrafos de texto são separados?	0,891	-3,03	24	No preenchimento de um formulário, é informado a forma de preenchimento?	1,043	-0,55
9	As palavras aparentemente clicáveis são de fato clicáveis?	1,075	0,16	25	Os dados obrigatórios são diferenciados dos dados opcionais de forma visualmente clara?	0,86	-1,04
10	Os títulos de telas, janelas e caixas de diálogo estão no alto, centrados ou justificados à esquerda?	1,376	-2,55	26	As mensagens de erro estão isentas de abreviaturas e/ou códigos gerados pelo sistema operacional?	1,073	-2,88
11	Todas as páginas possuem um campo de busca?	0,755	-2,70	27	Todas as páginas possuem os mesmos layouts e exibem ao usuário as mesmas características?	1,337	-1,58
12	Os resultados de busca permitem classificação por outros critérios além de custo?	1,204	0,46	28	O logotipo da empresa está no canto superior esquerdo em todas as páginas do site?	1,53	-1,74
13	Listas longas apresentam indicadores de continuação, de quantidade de itens e de páginas?	0,791	-0,28	29	Existe um link de um único clique que conduz a homepage?	1,178	-2,19
14	O preço de um produto consta ao lado da imagem ou do link do produto?	1,120	-2,63	30	Qualquer ação do usuário pode ser revertida através da opção DESFAZER?	1,35	-1,24
15	Existe uma orientação ao usuário quanto ao restante do site?	1,000	0,17	31	O site permite navegação em suas páginas em apenas uma janela?	1,056	-1,74
16	A maioria dos produtos possui informações sobre eles?	1,022	-2,11	32	Os links já visitados mudam de cor?	0,716	4,29

2003; WILSON, 2005). Geralmente a elaboração dos itens é realizada por especialistas ou extraída de literatura específica. Na presente aplicação, a elaboração dos itens foi embasada em referências específicas de usabilidade. Foram elaborados inicialmente 73 itens, todos com resposta dicotômica, do tipo “possui” ou “não possui” a característica em questão. A aplicação do conjunto de itens aos sites foi realizada pelos autores.

Para verificar a operacionalidade do conjunto bem como seu alinhamento com os pré-requisitos da TRI (unidimensionalidade, interdependência, clareza e objetividade), foi realizado um pré-teste, com cerca de 20 sites, quando foram eliminados 29 itens. Após este refinamento, os 44 itens restantes foram aplicados em 361 sites de *e-commerce* de origem brasileira. O tratamento dos dados foi feito no software BILOG-MG[®] produzido pela Scientific Software, Inc, Mislevy, R. J. e Bock, R. D, (1990). Através deste tratamento foi possível identificar os itens que apresentaram pouca informação para o conjunto, isto pôde ser verificado através do parâmetro de dificuldade do item. Desta forma, foram identificados 12 itens com baixa informação e que, portanto, não estavam medindo exclusivamente o grau de usabilidade dos sites. Os itens restantes com os seus parâmetros estimados podem ser visualizados do quadro 2.

É importante ressaltar que os valores referentes ao grau de dificuldade são referenciados a uma escala baseada na métrica (0,1), ou seja, média 0 (zero) e desvio-padrão 1 (um). Sendo assim, os valores, não possuem um significado prático em termos de usabilidade. Entretanto, possuem a capacidade de comparabilidade entre os itens e em relação a escala de usabilidade.

O parâmetro grau de dificuldade (b) não avalia a “qualidade” de um item, ele apenas indica qual a região, na escala proposta, em que o item possui maior informação. Já o grau de discriminação de um item (a) determina a “qualidade” do item, ou seja, quanto maior o valor de “a” maior o grau de discriminação na região de maior informação (“b”), conseqüentemente melhor o item. A figura 2 ilustra a Curva Característica do Item 30, que possui parâmetro de dificuldade -1,24 e parâmetro de discriminação igual a 1,35.

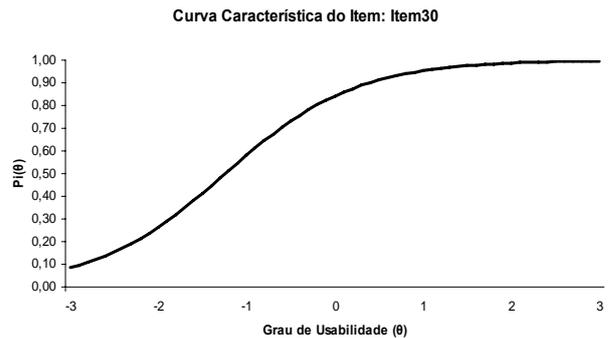


Figura 2: Curva Característica do Item 30

A CCI apresentada na figura 2, juntamente com seus parâmetros indica que um site com grau de usabilidade de -1,24 tem 50% de probabilidade de apresentar esta característica (qualquer ação do usuário pode ser revertida através da opção *desfazer*). Já um site com grau de usabilidade maior que 1,00 possui praticamente 100% de probabilidade de possuir esta característica. Considerando que a escala gerada possui uma média igual a zero e um desvio padrão igual a um, este item pode ser considerado “fácil”. Já o item 32, que possui parâmetro de discriminação igual a 4,29 é considerado mais difícil, ou seja, a maioria dos sites não modifica a cor dos links já visitado. Este tipo de conclusão pode ser estendido a todos os itens e uma vez que um site “responde” a este conjunto de itens seu grau de usabilidade pode ser estimado nesta escala também, permitindo, assim, a comparabilidade entre sites e a identificação das características mais relevantes.

Outra informação importante verificada neste estudo foi a convergência dos resultados estatísticos da TRI com os conceitos de usabilidade. Itens com baixo parâmetro de dificuldade, ou seja, mais básico, como os itens 3, 8, 10, 19, 20 estão diretamente relacionados com o design da página, ligados à exigência de memorização, legibilidade, padrões da web e modelos mentais. Estas características tendem a dificultar a interação do usuário com a página, entretanto, dificilmente o impedirão de realizar sua atividade. Todavia, itens com parâmetro de dificuldade mais alto como, por exemplo, os itens 9, 12, 15 e 18 estão relacionados com a navegabilidade do site, o que pode prejudicar significativamente a interação do usuário com o site.

A figura 3 apresenta a distribuição dos sites analisados, na escala gerada.

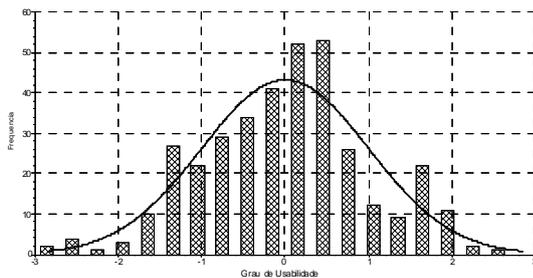


Figura 3: Distribuição do grau de usabilidade dos sites analisados

Observa-se que, embora a maioria dos itens apresentados possuam grau de dificuldade negativo na escala (ver quadro 2), a maioria dos sites analisados ficou em torno de zero. Com isto constata-se a independência da análise com a amostra, ou seja, na TRI, o parâmetro de dificuldade do item é independente de quem está respondendo o item, um ponto positivo em termos de confiabilidade e robustez do modelo gerado.

5. Considerações Finais

A Teoria da Resposta ao Item vem sendo amplamente utilizada nas áreas educacional e psicológica, sobretudo em países desenvolvidos. O presente artigo demonstra a viabilidade de sua utilização na avaliação de usabilidade. Uma vez que se trata de uma ferramenta estatística para criação de escalas padronizadas, com esta foi possível criar uma escala padronizada para medir o grau de usabilidade em sites de e-commerce. Isso possibilitou: (a) quantificar a relevância de cada aspecto de usabilidade avaliado, (b) identificar quais empresas possuem maior grau de usabilidade em relação às concorrentes, (c) identificar quais aspectos a empresa necessita desenvolver para melhorar sua usabilidade e (d) identificar claramente as limitações do instrumento, principalmente no que diz respeito a sua área de precisão.

As maiores desvantagens da TRI são a sua complexidade e a necessidade de amostras grandes, entretanto, suas vantagens compensam o esforço, uma vez que conferem confiabilidade, objetividade e robustez ao resultado obtido, características pouco verificadas nos métodos convencionais de avaliação de usabilidade.

4. Referências Bibliográficas

- ANDRADE, D. F.; TAVARES, H. R. e VALLE, R. C. *Teoria de Resposta ao Item: conceitos e aplicações*. ABE — Associação Brasileira de Estatística, 4º SINAPE, 2000.
- CYBIS, W. *Ergonomia e Usabilidade : conhecimentos, métodos e aplicações* / Walter Cybis, Adriana Holtz Betiol, Richard Faust. São Paulo : Novatec Editora, 2007.
- HAMBLETON, R. K. Emergence of Item Response Modeling in Instrument Development and Data Analysis. *Medical Care* v.38 n9 (Supplement II); p. 60-65, 2000.
- ISO 9241. *Ergonomic requirements for office work with visual display terminals (VDTs)*. Part 11 Guidance on usability; Draft International Standard ISO, 1998.
- JANDA, S., TROCCHIA, P.J. GWINNER, K. Consumer perceptions of internet retail service quality, *International Journal of Service Industry Management*, v. 13 n. 5, p. 412-431, 2002.
- KIM, S. STOEL, L. Dimensional hierarchy of retail website quality, *Information & Management*, v. 41 n. 5, p. 619-633, 2004.
- LIM, H. DUBINSKY, A.J. Consumers' perception of e-shopping characteristics: an expectancy-value approach, *Journal of Services Marketing*, v. 18 n.7, p. 500-513, 2004.
- LONG, M. McMELLON, C. Exploring the determinants of retail service quality on the internet, *Journal of Services Marketing*, v. 18 n. 1, p. 78-90, 2004.
- LORD, F. M. *A theory of test scores*. Psychometric Monograph No. 7, Psychometric Society, 1952.
- MISLEVY, D. J., BOCK, R. D. BILOG: *Item analysis and test scoring with binary logistic models* [Computer program]. Chicago: Scientific Software, 1990
- MORANDINI, M. *Ergo-Monitor: Monitoramento da Usabilidade em Ambiente Web por meio da Análise de Arquivos de Log* Tese Doutorado, Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis, SC, 2007.
- NIELSEN, J.; LORANGER, H. *Prioritizing Web Usability*. California : New Riders, California, 2006.



NIELSEN, J.; MARCK, Robert. *Usability Inspection Methods*. New York: John Wiley & Sons, 1994.

PASQUALI, L. *Psicometria: teoria dos testes na psicologia e na educação*. Petrópolis, RJ: Vozes, 2003.

TEZZA, R. *Proposta de um construto para medir usabilidade em site de e-commerce utilizando a Teoria da Resposta ao Item*. Dissertação de Mestrado - Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis, 2009.

WILSON, M. *Constructiong Measures: An Item Response Modeling Approach*, Lawrence Erlbaum Associates, Publishers Mahwah, New Jersey, London, 2005