

# Modelo multidimensional para mensurar qualidade em website de *e-commerce* utilizando a teoria da resposta ao item

## *Multidimensional model to measure quality in e-commerce websites using item response theory*

Rafael Tezza<sup>1</sup>  
Antonio Cezar Bornia<sup>2</sup>  
Dalton Francisco de Andrade<sup>3</sup>  
Pedro Alberto Barbetta<sup>3</sup>

**Resumo:** O presente artigo propõe um modelo multidimensional para medir a qualidade de websites comerciais com uso da Teoria da Resposta ao Item (TRI). A qualidade de um website engloba características técnicas (usabilidade/navegabilidade, apresentação da informação e interatividade) e não técnicas (*design*, estética, apelo visual, confiabilidade, hedonismo, imagem), configurando-se, teoricamente, como contexto multidimensional. As suposições iniciais das dimensões e elaboração dos itens foram baseadas em análise bibliográfica sobre o tema qualidade em websites de *e-commerce*. Um conjunto de 75 itens foi elaborado e uma amostra de 441 websites de *e-commerce* dos mais variados setores foi a ele submetido. O tratamento e análise dos dados foram feitos utilizando-se a TRI. Nessa etapa, foram discutidas questões referentes à dimensionalidade e à escolha do modelo mais adequado. Por fim, um modelo multidimensional com quatro dimensões foi ajustado contemplando as dimensões navegabilidade ou condução/orientação do usuário, acessibilidade e confiabilidade do sistema, interatividade e apresentação da informação.

**Palavras-chave:** Qualidade em websites; *E-commerce*; Modelagem; Teoria da Resposta ao Item Multidimensional.

**Abstract:** *The goal of this article is to propose a multidimensional modeling to measure the quality of commercial websites with the use of Item Response Theory (IRT). The quality of a website encompasses technical characteristics (usability-navigability, presentation of information and interactivity) and non-technical characteristics (design, aesthetics, visual appeal, reliability, hedonism, image), theoretically configuring a multidimensional context. The initial hypothesis of the dimensions and the elaboration of the items were based on a bibliographic analysis about the theme of e-commerce website quality. A set of 75 items was prepared and submitted to a sample of 441 e-commerce websites from a wide variety of sectors. The treatment and analysis of data was conducted using IRT. In this step, questions related to dimensionality and the choice of the most suitable model was discussed. Finally, a multidimensional model with four dimensions was adjusted.*

**Keywords:** *Quality in websites; E-commerce; Modeling; Multidimensional Item Response Theory.*

## 1 Introdução

O mercado de comércio eletrônico (*e-commerce*) vem crescendo exponencialmente desde o início da década passada em todo o mundo, aumentando seu poder de difusão e competitividade (Kim et al., 2012). Com isso, as empresas reconhecem cada vez mais a necessidade competitiva de estar presente na web e se engajar nesse mercado (Deng & Poole,

2012). Isso tem estimulado um crescimento maciço de sites comerciais e intensificado a concorrência por clientes *online*, gerando uma proliferação de sites fornecendo funcionalidades similares e produtos, serviços ou informações idênticas, permitindo que os clientes possam mudar de um site para outro sem esforço. Assim, para atrair e reter clientes por meio da

<sup>1</sup> Departamento de Administração Empresarial, Universidade do Estado de Santa Catarina – UDESC, Av. Madre Benvenuta, 2037, Itacorubi, CEP 88035-001, Florianópolis, SC, Brasil, e-mail: rafael.tezza@udesc.br

<sup>2</sup> Departamento de Engenharia de Produção, Universidade Federal de Santa Catarina – UFSC, Campus Universitário, Trindade, CP 476, CEP 88040-900, Florianópolis, SC, Brasil, e-mail: cezar@deps.ufsc.br

<sup>3</sup> Departamento de Informática e Estatística, Universidade Federal de Santa Catarina – UFSC, Campus Universitário, Trindade, CP 476, CEP 88040-900, Florianópolis, SC, Brasil, e-mail: dalton.andrade@ufsc.br, pedro.barbetta@ufsc.br

internet, uma organização precisa fazer seu website o mais simples e atraente, caso contrário os clientes potenciais podem simplesmente ir para outro website (Taylor & England, 2006).

Websites comerciais estão se tornando cada vez mais complexos, pois o número de funcionalidades oferecidas aos consumidores está aumentando constantemente, a fim de melhorar tanto o processo de coleta de informações quanto a experiência de compra *online* (Éthier et al., 2008). Com esses novos desenvolvimentos, gestores e desenvolvedores de websites estão sobrecarregados de abordagens com recomendações sobre como criar e manter um website atraente e de qualidade.

Pesquisadores de várias áreas desenvolveram e testaram instrumentos para medir a qualidade de website (por exemplo, Aladwani & Palvia, 2002; Loiacono et al., 2002; van der Merwe & Bekker, 2003). Nesses estudos, um grande número de itens foi gerado para medir a qualidade de websites e esses foram aplicados em uma variedade de sites comerciais ou não comerciais por meio de avaliação de usuários experientes, novatos, estudantes, desenvolvedores, entre outros, e os achados têm indicado que a qualidade em websites representa um construto multidimensional (Kim & Stoel, 2004; Hasan, 2016).

Nesse sentido, segundo Aladwani & Palvia (2002), a construção de medidas de qualidade no contexto de websites é uma tarefa desafiadora, por ser um conceito complexo e de natureza multidimensional. A maioria dos estudos envolve características técnicas tais como velocidade de *download* (Palmer, 2002; Galletta et al., 2004; Gata & Gilang, 2017), segurança (Flavián & Guinalíu, 2006; Braz et al., 2007), usabilidade (Bangor et al., 2008; Fang & Holsapple, 2007; Singh et al., 2016), qualidade do conteúdo (Agarwal & Venkatesh, 2002; Goldberg & Allen, 2008; Stoyanov et al., 2015) e características não técnicas, como estética (Pandir & Knight, 2006; Lindgaard et al., 2006; Stoyanov et al., 2015), confiabilidade (Seffah et al., 2006; Flavián et al., 2006) e apelo visual (Kulviwat et al., 2007; Lindgaard, 2007). Entretanto, poucos estudos têm explorado a hierarquia das dimensões desses elementos, de modo que pouco se sabe sobre as relações entre as diversas escalas ou a construção global de uma escala para medir qualidade do website. Uma alternativa para resolver esse problema é o desenvolvimento de uma sistemática ou modelo multidimensional considerando, da forma mais abrangente, os fatores que influenciam diretamente a qualidade de um website. A elaboração de um modelo em forma de uma medida multidimensional foi a proposta central deste artigo. Nesse sentido, a Teoria da Resposta ao Item Multidimensional (TRIM) surge como alternativa para o desenvolvimento desse processo, uma vez que está associada a uma estrutura matemática e

metodológica robusta, capaz de criar uma escala multidimensional na qual é possível posicionar na mesma escala item e respondente.

Este artigo é organizado da seguinte forma: após esta introdução é realizada uma breve definição de qualidade na web e criação de escalas, na sequência é apresentado o processo de medida proposto, seguido dos resultados e discussões, conclusões e referências.

## 2 Definição de qualidade na web

Reflexões a partir de estudos sobre a interação das pessoas com a tecnologia implicam que a avaliação dos clientes das novas tecnologias seja um processo distinto do tradicional (Parasuraman et al., 2005), pois envolve um modo diferente de fazer negócios, comumente chamado de *e-service*. Rust (2001) define *e-service* como o fornecimento de serviço utilizando redes eletrônicas.

Seguindo essa mesma ideia, Gefen (2002), Zeithaml et al. (2000), DeLone & McLean (2003), Silva et al. (2015) e outros têm se concentrado no desenvolvimento e adaptação de modelos para avaliação de serviço *online*. Entretanto, segundo Serkan et al. (2010), em comparação com a pesquisa abundante no contexto tradicional de qualidade em serviços, as pesquisas sobre a qualidade do serviço *online* ainda estão no começo.

Do ponto de vista técnico, as avaliações em serviços *online* receberam também influência dos estudos voltados para avaliação de sistemas de informação. Esses estudos atinham-se em conceituar tópicos específicos nesse contexto, tais como qualidade do gerenciamento de dados (Kaplan et al., 1998; Wang et al., 1995), qualidade da informação (King & Epstein, 1983; Haga & Zviran, 1994), qualidade de *softwares* (ISO, 1992; Schneidewind, 1992; Kitchenham & Pfllege, 1996), qualidade global do sistema (Kettinger & Lee, 1994; Nelson, 1996), entre outros. Esses estudos eram muito mais focados na avaliação do sistema, seu desempenho e sua relação com usuários específicos.

A partir de meados da década de 1990, com o desenvolvimento e popularização da internet, desenvolvedores e pesquisadores têm se esforçado para definir qualidade no contexto da internet (por exemplo: Barnes & Vidgen, 2000; Day, 1997; Lindroos, 1997; Xie et al., 1998; Loiacono et al., 2002). Lindroos (1997) discute as diferenças entre os sistemas de informação baseados na web e sistemas de informação convencionais a partir da perspectiva de qualidade de *software*. Barnes & Vidgen (2000), Loiacono et al. (2002), Parasuraman et al. (2005) e Ding et al. (2011) desenvolveram modelos focados em sites comerciais. Esses e vários outros estudos decompõem a qualidade dos websites em vários atributos. A criação de tais modelos é baseada principalmente em experiências de muitos anos

no desenvolvimento e manutenção de sistemas de informação e web. A validação desses modelos é feita principalmente por estudos empíricos, tais como a análise dos dados coletados em testes com usuários, questionários de satisfação e entrevistas.

Entretanto, a falta de uma definição clara dificulta a direção da pesquisa e impede a comparação e a integração dos resultados. Segundo Fassnacht & Koese (2006), dimensões e subdimensões propostas em um artigo podem não ser sempre comparáveis com as de outros estudos. Por exemplo, Gummerus et al. (2004) discutem a dimensão da qualidade da resposta em termos de *feedback* aos clientes, enquanto no trabalho de Loiacono et al. (2002) essa dimensão refere-se ao tempo de carregamento dos sites. Por meio de uma revisão da literatura no período de 2000 à 2016, tendo como referência as bases de dados Scopus, ScienceDirect, EBSCO e Emerald, usando as palavras-chave “*e-commerce*” ou “*web site*” ou “*website*” ou “*online shopping*” ou “*internet shopping*” e “*quality*” no título ou no resumo pode-se evidenciar o caráter multidimensional dos estudos. Dentre esses, encontraram-se abordagens práticas com no mínimo 3 e no máximo 15 dimensões. A maioria delas trabalhando com testes e percepção de usuários e utilizando como técnica de análise a análise fatorial exploratória e/ou confirmatória, demonstrando que o desenvolvimento e/ou aplicação de novas técnicas pode representar um avanço técnico-metodológico nesse campo. As dimensões mais encontradas nesses estudos foram apresentação da informação, segurança, usabilidade e interatividade.

Além disso, a grande maioria dos estudos encontrados é na realidade adaptação ou aplicação de modelos já existentes e, de acordo com Fassnacht & Koese (2006), a estrutura e o significado das dimensões elaboradas têm sido geralmente determinados a posteriori pelos resultados de análise de dados.

Uma questão relevante dos estudos nessa área, do ponto de vista metodológico, é a grande diferença no tamanho da amostra. Por exemplo, enquanto Jun et al. (2004) utilizam 137 usuários (estudantes e profissionais), Yang et al. (2005) utilizam 1.992 usuários. Como a grande maioria das abordagens pesquisadas utiliza como procedimento para análise dos dados a análise fatorial, pode-se comparar a confiabilidade das escalas propostas. No caso da escala proposta por Jun et al. (2004), a confiabilidade (alpha de Cronbach) ficou entre 0,59 e 0,92, enquanto que no estudo de Yang et al. (2005) variou de 0,66 a 0,89. Geralmente assume-se como aceitável valores acima de 0,70, mas, segundo Kline (2000), em casos com grande diversidade de construtos como, por exemplo, testes psicológicos, são aceitáveis valores abaixo de 0,70. A maioria dos estudos pesquisados apresentou um alpha de Cronbach maior que 0,70, com exceção do estudo de

Ibrahim et al. (2006), o qual apresenta uma amplitude de 0,33 a 0,84.

O presente estudo se encaixa na classificação de Cristobal et al. (2007) como um estudo da qualidade e design de website. Dentro desse escopo, assume-se qualidade em websites como a qualidade de um sistema de informação que, segundo Loiacono et al. (2002), foca-se no armazenamento, processamento, apresentação e transferência da informação.

Sendo assim, o conceito adotado de qualidade em websites é o de um conjunto de características técnicas e não técnicas de um sistema web, as quais permitem ao usuário realizar seus objetivos em um website de forma acessível, eficiente e agradável. Como características técnicas entende-se usabilidade/navegabilidade, apresentação da informação, acessibilidade e interatividade do sistema. Como características não técnicas entende-se *design*, estética, apelo visual/comercial, confiabilidade, hedonismo, empatia.

### 3 Criação de escalas: teoria da resposta ao item e qualidade na web

A construção de escalas de medida facilita o entendimento de conceitos complexos. Três razões explicam isso. Primeiramente, procurar por vários aspectos de uma variável fornece a criação de conhecimento sobre ela. Segundo, várias perspectivas realçam o intervalo de variação, o que permite distinções mais precisas, principalmente se envolver medidas ordinais. Em terceiro lugar, a construção de medidas permite uma redução eficiente dos dados por meio da qual uma pontuação numérica pode representar a posição ordinal em que determinada característica do item ou de elementos da população em avaliação se encontram na escala, o que permite comparabilidade (Babbie, 2005).

Para construir uma escala baseada num conjunto de itens, as variáveis devem estar ligadas teórica ou conceitualmente com o que a pesquisa visa medir (Baker, 2009). Com isso, a maior implicação das escalas é a possibilidade de comparação entre escores e estatísticas apropriadas que resumem esses escores. Em uma escala de medida, o papel dos números é representar um objeto, ou uma pessoa, de forma que a relação entre os números represente a relação empírica entre os objetos ou entre as pessoas. Essa representatividade fornece ao objeto importantes propriedades tais como o poder de distinção entre objetos, ordem, adição e razão (Embretson & Reise, 2000).

A Teoria de Resposta ao Item (TRI) é composta de um conjunto de modelos probabilísticos que relaciona um traço latente de um respondente ( $\Theta$ ), que não pode ser medido diretamente, com a probabilidade de

ele responder a um item dentro de uma determinada categoria (Lord, 1980).

Na TRI, a escolha do modelo matemático depende basicamente do tipo de item e representa a probabilidade de resposta a um item em função dos parâmetros do item e da proficiência do respondente (Tavares et al., 2004; Reise et al., 1993). O modelo mais utilizado para itens com resposta dicotômica é o modelo logístico de dois parâmetros unidimensional (ML2P) desenvolvido por Birnbaum (1968), com base em Lord (1952). Sendo  $U_{ij} = 1$  a probabilidade de o indivíduo  $j$  responder afirmativamente ao item  $i$ , o ML2P é representado pela Equação 1:

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

em que  $\theta$  é o traço latente que se quer medir, supostamente, numa escala com média zero e desvio padrão um;  $\theta_j$  é o valor do traço latente para o indivíduo  $j$ ;  $b_i$  é o valor do traço latente em que a probabilidade do item  $i$  ser respondido afirmativamente é igual a 0,5;  $a_i$  é o parâmetro de discriminação do item  $i$ , proporcional à inclinação da Curva Característica do Item (CCI) no ponto  $b_i$ . A CCI representa a relação entre a resposta prevista ao item e o traço latente do indivíduo (Reckase, 1997).

Uma das suposições empregadas nesse modelo é a da unidimensionalidade, a qual define que todos os itens do teste estão medindo o mesmo traço latente ou a mesma composição de múltiplas habilidades. Entretanto, existem muitas situações em que os itens que compõem o instrumento de medida podem estar medindo diferentes dimensões do traço latente, ou diferentes composições de múltiplas habilidades (Ackerman, 1994) como, por exemplo, a qualidade de websites comerciais.

Reckase (1985) descreve o modelo multidimensional compensatório de dois parâmetros como uma extensão multidimensional do ML2P apresentado na Equação 1, da forma apresentada na Equação 2:

$$P(U_{ij} = 1 | \theta'_{jk}, d_i, a'_{ik}) = \frac{e^{(\sum_{k=1}^m a_{ik} \theta_k + d_i)}}{1 + e^{(\sum_{k=1}^m a_{ik} \theta_k + d_i)}}, \quad (2)$$

em que:  $U_{ij}$  = resposta da pessoa  $j$  ao item  $i$  (0 ou 1);  $a_{ik}$  = parâmetro de discriminação do item  $i$  na dimensão  $k$ ;  $\theta_k$  = traço latente da pessoa  $j$  na dimensão  $k$ ;  $d_i$  = parâmetro de dificuldade do item  $i$ .

O expoente de  $e$  no modelo 2 pode ser escrito conforme Equação 3:

$$\sum_{k=1}^m a_{ik} \theta_k + d_i = a_{i1} \theta_{j1} + a_{i2} \theta_{j2} + \dots + a_{im} \theta_{jm} + d_i \quad (3)$$

A Equação 3 mostra que o expoente é uma função linear de elementos de  $\theta$ , com o parâmetro  $d$  como a ordenada na origem e os elementos do vetor  $a$  como parâmetros de inclinação ou de discriminação.

Uma das propriedades desse modelo é que a expressão representada no expoente define uma reta num espaço de  $k$  dimensões que pode gerar linhas de equiprobabilidade, ou seja, essa forma multidimensional permite que existam infinitas combinações lineares que resultam no mesmo expoente, gerando a mesma probabilidade de acerto. Essa propriedade confere ao modelo a característica compensatória. Para mais detalhes ver Reckase (2009).

A utilização dos modelos da teoria da resposta ao item multidimensional (TRIM) para lidar com problemas de medição em avaliação educacional em larga escala tem sido realizada desde o início dos anos 1990 (Ackerman, 1992; Camilli, 1992; Embretson, 1991; Glas, 1992; Oshima & Miller, 1992; Reckase & McKinley, 1991). No entanto, segundo Adams et al. (1997), Hartig & Höhler (2008) e Rauch & Hartig (2010), a aplicação de modelos em teste prático fora da área educacional é relativamente rara. Na avaliação de sistemas, mais especificamente sistemas web, as aplicações da TRI são muito raras, Schmettow & Vietze (2008) propõem a aplicação do modelo de Rasch (Rasch, 1960) para o processo de inspeção de usabilidade, Tezza et al. (2011) aplicaram o modelo logístico de dois parâmetros unidimensional para mensurar usabilidade em websites comerciais e Tezza et al. (2016) comparam modelos hierárquicos e não hierárquicos na TRI no contexto de websites comerciais. No entanto, não foi encontrada na literatura a apresentação de um modelo multidimensional da TRI interpretável para avaliação da qualidade de websites comerciais.

## 4 Método

A construção dos itens partiu de um levantamento da literatura conforme descrito na seção 2. Procurou-se consolidar os diversos estudos presentes na literatura específica e desenvolver um instrumento abrangente e consistente. Mais especificamente, a construção dos itens se deu por meio da associação dos conceitos resultantes da análise de 213 artigos e livros pesquisados. Alguns itens foram elaborados a partir dos conceitos mais recorrentes na literatura. Outros foram extraídos integralmente de estudos anteriores como, por exemplo, Tezza et al. (2011) e W3C (2008).

O planejamento do levantamento de dados seguiu a proposta de Stepchenkova et al. (2010) e Tezza et al. (2011), no qual os itens são construídos para avaliar o sistema e não com o objetivo de captar percepção de clientes ou usuários, tendo, portanto, a característica objetiva de um *checklist*. Com isso, trabalhou-se a elaboração de itens objetivos, capazes de verificar a existência ou não de determinada característica associada à qualidade do website comercial. Sendo assim, todos os itens foram formulados de forma objetiva e com respostas dicotômicas.

Foram elaborados 75 itens. Após a elaboração dos itens os mesmos foram submetidos à avaliação de três especialistas, com o objetivo de verificar a adequação dos itens ao constructo – qualidade em websites comerciais.

Os dados utilizados na presente análise foram formados por uma amostra da população de sites de *e-commerce* de origem brasileira. A coleta dessa amostra foi feita com o auxílio de buscadores como Google.com, Bing.com e Yahoo.com no período de janeiro e fevereiro de 2012. Foram analisados 441 sites de *e-commerce*, abrangendo vários gêneros, escolhidos de forma aleatória. Durante o processo de amostragem tomou-se o cuidado de garantir que a mesma fosse a mais diversificada possível. Para isso, além de sites com os mais diversos gêneros de produtos comercializados, observou-se a variedade de estilos de *design*, contemplando desde sites simples até muito elaborados, o que não necessariamente implica em maior qualidade mas contribui para a diversidade, necessária para aplicação da Teoria da Resposta ao Item (Tezza et al., 2011).

Dos 75 itens gerados, 56 deles tiveram suas respostas obtidas por meio de coleta manual, na qual o pesquisador, após navegação no website em análise, responde se o mesmo possui ou não as características em questão. Esse método manual de coleta de dados em websites baseou-se em Pinterits et al. (2006), Al-Khalifa (2010), Stepchenkova et al. (2010) e Tezza et al. (2011). A coleta manual foi realizada pelo primeiro autor. Essa coleta manual levou dois meses – março a maio de 2012. Ao todo foram avaliados 441 websites. Os 19 itens restantes foram analisados de forma semiautomática utilizando a ferramenta livre AChecker (2012). Essa ferramenta verifica as diretrizes de acessibilidade, com o objetivo de estabelecer o nível de acessibilidade de websites ao detectar erros de acordo com as diretrizes de Acessibilidade para Conteúdo Web (WCAG) 2.0.

Constituiu o ambiente para coleta de dados um computador com tela de 15". O navegador utilizado foi o Internet Explorer versão 7. A resolução da tela utilizada foi de 1.024 por 768 *pixels*.

Na etapa de análise dos dados foi realizada inicialmente uma análise fatorial exploratória para verificar a qualidade dos itens e, posteriormente, análises de dimensionalidade, também por meio

de uma análise exploratória para identificação do número de dimensões e a adequação de cada item às dimensões. Por fim foi verificada a adequação do modelo multidimensional compensatório da teoria da resposta ao item e interpretação da escala.

## 5 Resultados e discussões

### 5.1 Avaliação da qualidade do conjunto de itens iniciais

Inicialmente foi realizada uma avaliação da qualidade do conjunto inicial de dados com 75 itens, por meio do método de análise fatorial exploratória de informação plena. A análise foi conduzida baseada na verificação das cargas fatoriais de cada item, o que reflete a relação do item com os fatores subjacentes (traços latentes) presentes no conjunto de dados.

Para verificação das cargas fatoriais, utilizou-se, após analisar estruturas com 1, 2, 3 e 4 dimensões, a estrutura de 3 dimensões. A utilização de 3 dimensões se deu com base no critério sugerido por Chalmers (2012), o qual afirma que o número de dimensões que gera um melhor ajuste aos dados pode ser verificado por meio de uma comparação de modelos, utilizando uma análise de variância genérica (ANOVA) implementada no *software* R com base no método da análise fatorial de informação plena. Como resultado tem-se a estatística qui-quadrado ( $\chi^2$ ) associada ao teste de razão de verossimilhanças, bem como a diferença no AIC (Akaike Information Criterion – AIC) e BIC (Bayesian Information Criterion – BIC), que são estatísticas de comparação de ajuste de modelos (quanto menor, melhor). Foi feita a comparação de quatro modelos, o primeiro assumindo uma dimensão (Mod1), o segundo, duas (Mod2), o terceiro, três (Mod3) e o quarto, quatro (Mod4). A Tabela 1 demonstra os resultados.

Verifica-se, na Tabela 1, que a diferença entre o modelo 1 e o modelo 2 é estatisticamente significativa para  $\alpha = 0,05$ , (hipótese alternativa aceita), indicando que o modelo que assume duas dimensões possui mais informação que o modelo de uma dimensão, e isso também pode ser visualizado nos critérios de informação AIC e BIC, nos quais verifica-se que existem diferenças positivas do modelo 1 em relação ao modelo 2. O mesmo ocorre na comparação do modelo 2 e do modelo 3. No entanto, nesse caso,

**Tabela 1.** Comparação dos modelos de 1, 2, 3 e 4 dimensões.

Modelos	Diferença qui-quadrado			AIC	BIC
	$\chi^2$	Graus de liberdade	p-valor		
Mod1 x Mod2	525,50	74	<0,01	377,50	74,91
Mod2 x Mod3	283,32	73	<0,01	137,32	-161,18
Mod3 x Mod4	83,88	72	0,16	-60,12	-354,53

AIC - Akaike Information Criterion; BIC - Bayesian Information Criterion.

verifica-se no AIC que há diferença positiva entre os modelos 2 e 3, mas no BIC o mesmo não ocorre.

Usando análise fatorial de informação plena, verificou-se que a maioria dos itens apresentou cargas fatoriais superiores a 0,3 em alguma dimensão, o que é considerado por alguns autores (Johnson & Wichern, 2007; Hair et al., 2009) valor mínimo para se considerar o item na interpretação da dimensão. Itens com carga fatorial inferior a 0,3 em todas as dimensões foram assumidos como itens pouco informativos para o construto, sendo assim, foram eliminados. Essa ação resultou na retirada de 18 itens. A nova análise do conjunto, após essa primeira retirada, mostrou mais dois itens com carga fatorial, em módulo, inferior a 0,3 em todas as dimensões: os itens 1 e 42, os quais também foram eliminados. Foram realizadas cinco novas análises até extrair-se todos os itens remanescentes com carga fatorial inferior a 0,3. Desta forma foram eliminados ao todo 31 itens, restando 44 nessa etapa.

## 5.2 Avaliação das dimensões

Após a verificação da qualidade do conjunto inicial de 75 itens por meio da análise das cargas fatoriais partiu-se para a avaliação da dimensionalidade dos 44 itens restantes. Para avaliar a dimensionalidade do construto utilizaram-se o método de análise de componentes principais e o método de análise fatorial de informação plena. No primeiro método, o número de dimensões foi verificado com base na matriz de correlação tetracórica e na análise paralela. Para isso foi utilizado o pacote *psych* (Revelle, 2012) implementado no *software R* (R Core Team, 2012). A análise de componentes principais, realizada por meio da matriz de correlação tetracórica, sugeriu a existência de 5 dimensões, assumindo-se o critério de pelo menos 50% da variância comum para o conjunto de fatores. A Tabela 2 mostra os primeiros 10 autovalores considerando os 44 itens.

Alguns autores, como McDonald (1981) e Spector et al. (1997), argumentam que a análise fatorial clássica tende a superestimar o número de dimensões, criando fatores espúrios quando não há uniformidade na dificuldade dos itens. Dessa forma, não é muito conveniente considerar um número muito alto de dimensões tendo em vista também o número limitado da amostra.

Já a análise fatorial feita pela análise fatorial de informação plena mostrou que a primeira dimensão explica apenas 10,41% da variação total do conjunto de 44 itens, o que sugere um construto não unidimensional. A comparação do ajuste de modelos TRIM de 1, 2, 3 e 4 dimensões sugere 3 dimensões, pelo critério AIC, como pode ser visualizado na Tabela 3.

Complementarmente, foi usado o critério empírico, recomendado por diferentes autores, como Ford et al. (1986), Fleck & Bourdel (1998), Podsakoff et al. (1997). Esse método propõe considerar as dimensões que possam ser bem interpretadas. Nesse contexto, verificou-se melhor interpretação para o modelo TRIM com quatro dimensões.

A Tabela 4 mostra a descrição dos 44 itens e a distribuição das cargas fatoriais de cada item nas 4 dimensões, utilizando a rotação ortogonal varimax. A primeira dimensão agrupa itens com conteúdo voltado para acessibilidade e utilização do sistema, ou seja, verificação de possíveis entraves no sistema, que possam impedir, confundir ou dificultar o acesso dos usuários, particularmente usuários que possuam alguma limitação física ou cognitiva. Nota-se que a maioria dos itens dessa dimensão pertence ao conjunto de recomendações de acessibilidade WCAG 2.0 (W3C, 2008). A segunda dimensão traz itens relacionados diretamente com a apresentação da informação, praticamente todos os itens com maior carga fatorial nessa dimensão remetem a algum conceito de organização ou apresentação da informação. A terceira dimensão, assim como a

**Tabela 2.** Valores próprios da matriz de correlação tetracórica dos primeiros 10 autovalores do conjunto de 44 itens.

Dimensão	1	2	3	4	5	6	7	8	9	10
Autovalor	8,95	4,44	3,89	3,15	2,55	2,12	1,79	1,61	1,60	1,47
Proporção acumulada da variação explicada	20,34	30,42	39,26	46,41	<b>52,22</b>	57,03	61,09	64,74	68,38	71,72

**Tabela 3.** Seleção do melhor modelo com base nos critérios de informação AIC e BIC.

Modelo	-2 Log da máxima verossimilhança	AIC	BIC
Mod1	-6849,03	14068,05	14824,53
Mod2	-6617,70	13691,87	14623,69
Mod3	-6531,43	<b>13602,87</b>	14706,91
Mod4	-6493,57	13609,14	14880,84

AIC - Akaike Information Criterion; BIC - Bayesian Information Criterion.

**Tabela 4.** Análise fatorial utilizando o método da informação plena, rotação ortogonal varimax aos 44 itens em 4 dimensões.

Item	Descrição	F_1	F_2	F_3	F_4	Comunalidade
3	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?	0,130	0,660	0,030	-0,020	0,454
5	O site possui imagens em movimento que podem distrair o usuário?	-0,160	0,140	-0,200	-0,440	0,279
6	Os links promocionais levam diretamente para a oferta anunciada?	0,390	0,680	0,110	-0,090	0,635
8	O site possui menu em cascata?	-0,060	0,070	-0,190	-0,570	0,370
10	Existe agrupamento de subcategorias?	-0,050	0,300	0,030	0,530	0,374
12	Existem informações de contato telefônico ou endereço?	-0,060	0,450	0,380	-0,300	0,441
19	Palavras aparentemente clicáveis são de fato clicáveis?	0,290	0,390	0,220	0,400	0,445
21	Todas as páginas possuem um campo de busca?	0,450	0,280	0,590	0,300	0,719
22	Quando há rolagem, não existem elementos de design (na tela inicial) que pareçam com marcadores de final de página?	0,260	0,380	0,270	-0,280	0,363
23	O logotipo da empresa está no canto superior esquerdo em todas as páginas do site?	0,240	0,480	0,260	0,270	0,429
24	Existe um link de um único clique que conduz a homepage?	0,520	0,250	0,750	0,320	0,998
25	O site permite navegação em suas páginas em apenas uma janela, ou seja, não há abertura de novas janelas em meio a navegação?	0,110	-0,030	0,820	-0,280	0,764
27	Existe uma lista de perguntas frequentes - FAQs?	0,030	0,300	-0,080	0,480	0,328
28	Ao digitar uma pesquisa no campo de busca, o buscador vai dando sugestões?	0,060	0,020	0,180	0,660	0,472
29	O sistema de busca é flexível quanto aos termos utilizados pelo usuário, ou seja, caso o usuário insira algum termo com grafia errada, o sistema de busca sugere correção?	0,220	0,170	-0,170	0,530	0,387
30	Os resultados de busca permitem classificação por outros critérios além de custo?	0,290	0,390	0,300	0,500	0,576
32	Ícones de continuação de página estão visíveis?	0,290	0,310	0,310	0,290	0,360
33	O preço de um produto consta ao lado da imagem ou do link do produto?	0,280	0,710	0,380	0,000	0,727
35	Em produtos em que existam mais de uma perspectiva, é possível visualizar todas as perspectivas?	0,050	0,330	0,050	0,330	0,223
37	Os produtos possuem informações suficientes (tamanho, características básicas)?	0,270	0,620	-0,040	0,110	0,471
38	Existe a possibilidade de inserir feedback sobre os produtos por parte do consumidor?	0,400	0,200	0,290	0,460	0,496
40	O site apresenta produtos relacionados no final da página?	0,070	0,450	0,030	0,240	0,266
43	O site possui multimídia para apresentação de produtos?	0,000	0,520	-0,090	0,080	0,285

F\_1 – Dimensão Acessibilidade/confiabilidade do sistema; F\_2 – Dimensão Apresentação da informação; F\_3 – Dimensão Controle por parte do usuário/interatividade; F\_4 – Dimensão Orientação do usuário/navegabilidade.

**Tabela 4.** Continuação...

Item	Descrição	F_1	F_2	F_3	F_4	Comunalidade
45	Existe indicação de ambiente seguro no momento da efetivação da compra?	0,130	0,520	0,020	0,330	0,397
46	No preenchimento de formulários, o usuário consegue visualizar os próximos passos na interface?	0,210	0,200	0,530	0,410	0,533
47	O site possui outras formas de pagamento além do cartão de crédito?	0,240	0,430	0,470	-0,240	0,521
48	É possível saber os custos totais antes de fazer cadastro (inclusive frete)?	0,330	0,600	0,290	0,250	0,616
52	Os dados obrigatórios são diferenciados dos dados opcionais de forma visualmente clara?	0,280	0,220	0,440	0,220	0,369
55	As mensagens de erro estão isentas de abreviaturas e/ou códigos gerados pelo sistema operacional?	0,210	0,100	0,540	-0,200	0,386
56	Qualquer ação do usuário pode ser revertida através da opção DESFAZER ou VOLTAR?	0,580	0,550	0,280	0,100	0,727
57	Todo o conteúdo não textual que é apresentado ao usuário tem uma alternativa em texto que serve a propósito equivalente?	-0,710	-0,050	0,400	-0,010	0,667
58	As informações, a estrutura e as relações transmitidas através de apresentação podem ser determinadas de forma programática ou estão disponíveis no texto?	-0,960	0,160	0,130	0,190	1,000
59	A cor não é utilizada como o único meio visual de transmitir informações, indicar uma ação, pedir uma resposta ou distinguir um elemento visual?	0,150	-0,460	0,520	0,270	0,577
60	A apresentação visual de texto e imagens de texto tem uma relação de contraste de, no mínimo, 4,5:1?	-0,720	-0,120	0,010	0,510	0,793
61	Exceto para legendas e imagens de texto, o texto pode ser redimensionado sem tecnologia assistiva até 200% sem perder conteúdo ou funcionalidade?	-0,340	-0,180	0,430	0,280	0,411
64	Para cada limite de tempo definido pelo conteúdo, no mínimo uma das seguintes afirmações é verdadeira: - Desligar: O usuário pode desligar o limite de tempo antes de o atingir; ou - Ajustar: O usuário pode ajustar o limite de tempo antes de o atingir, acima de um grande intervalo que dure, no mínimo, dez vezes mais do que a predefinição; ou - Prolongar: O usuário é avisado antes de o tempo expirar e tem, no mínimo, 20 segundos para prolongar o limite de tempo com uma ação simples (por exemplo, “pressionar a barra de espaços”), e o usuário pode prolongar o limite de tempo, no mínimo, dez vezes; ou - Exceção em Tempo Real: O limite de tempo é uma parte necessária de um evento em tempo real (por exemplo, um leilão), e não é possível nenhuma alternativa ao limite de tempo; ou - Exceção Essencial: O limite de tempo é essencial e prolongá-lo iria invalidar a atividade; ou - Exceção de 20 Horas: O limite de tempo é superior a 20 horas.	0,220	-0,600	0,550	0,100	0,721

F\_1 – Dimensão Acessibilidade/confiabilidade do sistema; F\_2 – Dimensão Apresentação da informação; F\_3 – Dimensão Controle por parte do usuário/interatividade; F\_4 – Dimensão Orientação do usuário/navegabilidade.



Tabela 4. Continuação...

Item	Descrição	F_1	F_2	F_3	F_4	Comunalidade
65	Para informações em movimento, em modo intermitente, em deslocamento ou em atualização automática, todas as seguintes afirmações são verdadeiras: <ul style="list-style-type: none"> <li>- Em movimento, em modo intermitente, em deslocamento: Para quaisquer informações em movimento, em modo intermitente ou em deslocamento, que (1) sejam iniciadas automaticamente, (2) durem mais de cinco segundos, e (3) sejam apresentadas em paralelo com outro conteúdo existe um mecanismo para o usuário colocar em pausa, parar, ou ocultar as mesmas, a menos que o movimento, o modo intermitente ou o deslocamento façam parte de uma atividade na qual sejam essenciais; e</li> <li>- Em atualização automática: Para quaisquer informações em atualização automática que (1) sejam iniciadas automaticamente e (2) sejam apresentadas em paralelo com outro conteúdo, existe um mecanismo para o usuário colocar em pausa, parar ou ocultar as mesmas, ou controlar a frequência da atualização, a menos que a atualização automática faça parte de uma atividade na qual é essencial.</li> </ul>	0,060	-0,590	0,670	0,230	0,854
66	Está disponível um mecanismo para ignorar blocos de conteúdo que são repetidos em várias páginas web?	0,440	-0,060	0,320	0,400	0,460
69	Os cabeçalhos e as etiquetas descrevem o tópico ou a finalidade?	-0,430	0,160	-0,230	-0,200	0,303
70	A linguagem humana predefinida de cada página web pode ser determinada de forma programática?	-0,180	-0,050	0,420	0,220	0,260
71	Alterar a definição de um componente de interface de usuário não provoca, automaticamente, uma alteração de contexto, a menos que o usuário tenha sido avisado sobre essa situação antes de utilizar o componente?	-0,490	0,220	0,220	0,120	0,351
73	Se um erro de entrada for automaticamente detectado, o item que apresenta erro é identificado e o erro é descrito ao usuário por texto?	0,180	-0,740	0,520	-0,050	0,853
74	Etiquetas ou instruções são fornecidas quando o conteúdo exige a entrada de dados por parte do usuário?	-0,880	0,120	0,260	0,130	0,873
75	No conteúdo implementado utilizando linguagens de marcação, os elementos dispõem de marcas de início e de fim completas, os elementos estão encaixados de acordo com as respectivas especificações, os elementos não contêm atributos duplicados, e todos os IDs são exclusivos, exceto quando as especificações permitem essas características?	-0,410	0,150	0,170	-0,120	0,234

F\_1 – Dimensão Acessibilidade/confiabilidade do sistema; F\_2 – Dimensão Apresentação da informação; F\_3 – Dimensão Controle por parte do usuário/interatividade; F\_4 – Dimensão Orientação do usuário/navegabilidade.

segunda, apresenta todos os itens com maior carga nessa dimensão no mesmo sentido, ou seja, todos positivos, com exceção do item 73. Analisando-se os conceitos envolvidos em cada item característico dessa dimensão, verifica-se que a maioria possui conceito adjacente ao controle por parte do usuário ou interação desse com o sistema. Na quarta dimensão, os itens 5 e 8 apresentaram carga em sentido contrário ao da maioria. Essa peculiaridade pode ser justificada pela própria característica dos itens nesse contexto, ou seja, considerando-se que todos os itens dessa dimensão estão relacionados diretamente com o conceito de orientação do usuário durante a navegação. O Quadro 1 mostra um resumo das dimensões e os respectivos itens de maior carregamento nessas, segundo a análise fatorial.

### 5.3 Aplicação do modelo multidimensional compensatório da TRI

Dadas as conclusões e verificações discutidas na seção anterior, assumiu-se que o construto contendo 44 itens é composto de 4 dimensões. Sendo assim, trabalhou-se uma modelagem multidimensional de 4 dimensões, utilizando-se a teoria da resposta ao item. A TRIM (Teoria da Resposta ao Item Multidimensional) possui algumas similaridades com a análise fatorial, entretanto a TRIM apresenta vantagens consideráveis ante a abordagem puramente fatorial, principalmente pelo fato de tratar os itens de uma forma individual e não apenas em termos de agrupamento fatorial. Nessa linha é possível analisar, além da qualidade do item, seu grau de dificuldade e, com isso, estimar escores aos respondentes. Segundo Bartolucci et al. (2012), nenhuma outra abordagem estatística permite estudar a dimensionalidade e o poder de discriminação de itens de maneira mais simples do que a TRIM.

A estimação dos parâmetros multidimensionais dos 44 itens apresentou alguns problemas pontuais com quatro itens (itens 5, 24, 58, 73). Tais itens apresentaram parâmetros superestimados e/ou com erro padrão alto e, portanto, foram eliminados.

A Tabela 5 mostra a estimativa dos parâmetros dos 40 itens utilizando o modelo multidimensional compensatório de dois parâmetros (Equação 2) por meio do *software* flexMIRT™ (Cai, 2012). A Tabela 5

mostra também os parâmetros de discriminação de cada item em cada dimensão ( $a_1, a_2, a_3, a_4$ ) e o parâmetro de discriminação multidimensional (MDISC). Quanto maior o parâmetro de discriminação multidimensional (MDISC), maior é o poder de discriminação multidimensional do item, ou seja, mais o item diferencia o nível de qualidade para o website. Além disso, a Tabela 5 mostra o valor do parâmetro de localização  $d$  relativa a cada item, que, segundo Reckase (1997), está relacionado à dificuldade do item.

O parâmetro de discriminação, assim como a carga fatorial, pode ser arbitrário em cada dimensão, considerando que possa haver algum tipo de rotação, esses carregamentos podem ser comparados identificando-se itens semelhantes na mesma dimensão, da mesma forma como é feita a análise fatorial (Reckase, 1997). Sendo assim, espera-se que o parâmetro de discriminação de um determinado item seja maior na dimensão em que ele tenha maior poder de discriminação, ou seja, na dimensão em que possua mais informação.

A estrutura apresentada na Tabela 5 assemelha-se à apresentada na análise fatorial da Tabela 4, a qual considera 4 dimensões. Entretanto, a configuração multidimensional da Tabela 5 apresenta algumas peculiaridades que a diferenciam daquela. Primeiramente, quatro itens foram retirados por apresentarem problemas nas estimativas dos parâmetros. Em segundo, apesar de possuírem uma base estatística comum, análise fatorial de informação plena e a teoria de resposta ao item multidimensional, a última trabalha com estimação de mais parâmetros, o que justifica algumas particularidades no processo de estimação, as quais podem gerar problemas de convergência como os que ocorreram nos itens 5, 24, 58 e 73. Por fim, a consequência gerada pelas reestimativas dos parâmetros é a possibilidade de reorganização dos itens nas dimensões e de alteração no significado base de alguma das dimensões, o que de fato ocorreu, particularmente com a dimensão caracterizada na análise fatorial como acessibilidade e utilização do sistema e agora na reestimação da TRIM, como acessibilidade e confiabilidade do sistema, representada pela dimensão 2.

Os valores em negrito na Tabela 5 destacam a dimensão na qual o item possui maior parâmetro de discriminação, indicando em qual dimensão o item

**Quadro 1.** Classificação dos itens conforme a dimensão, segundo a análise fatorial.

Dimensão	Itens
Acessibilidade/confiabilidade do sistema	56*, 57, 58, 60, 66*, 69, 71, 74, 75
Apresentação da informação	03, 06, 12, 22, 23, 32, 33, 35, 37, 40, 43, 45, 48, 56, 64*, 73*
Controle por parte do usuário/interatividade	21, 24, 25, 32, 46, 47, 52, 55, 59, 61, 64, 65, 70
Orientação do usuário/navegabilidade	05*, 08*, 10, 19, 27, 28, 29, 30, 35, 38, 66

\*Itens com carregamento contrário à maioria dos itens da dimensão.

Tabela 5. Estimativas dos parâmetros multidimensionais para os 40 itens, assumindo o modelo multidimensional de dois parâmetros compensatório.

Item	a 1	s.c.	a 2	s.c.	a 3	s.c.	a 4	s.c.	MDISC	d	s.c.	MDIFF
3	<b>1,43</b>	0,78	0,65	0,76	-0,34	0,64	0,66	0,74	1,74	4,23	0,82	-2,43
6	<b>2,29</b>	1,18	0,98	0,89	-0,03	0,70	0,39	0,93	2,52	4,88	1,39	-1,94
8	-0,02	0,35	0,29	0,40	<b>-0,99</b>	0,33	-0,70	0,39	1,25	0,79	0,19	-0,63
10	0,51	0,42	-0,15	0,43	0,39	0,42	<b>1,20</b>	0,46	1,37	2,09	0,36	-1,53
12	0,66	0,55	<b>1,47</b>	0,58	-0,06	0,57	-0,02	0,50	1,61	2,53	0,44	-1,57
19	<b>1,20</b>	0,39	0,14	0,45	0,75	0,38	0,92	0,40	1,69	1,50	0,25	-0,89
21	1,73	1,03	1,26	1,00	<b>2,35</b>	1,24	0,74	0,99	3,26	5,57	1,46	-1,71
22	<b>1,11</b>	0,79	0,94	0,71	0,33	0,63	-0,48	0,67	1,57	4,02	0,90	-2,57
23	<b>1,22</b>	0,50	0,42	0,62	0,44	0,59	0,91	0,63	1,64	3,61	0,52	-2,20
25	0,27	0,93	<b>2,39</b>	1,30	1,54	0,91	-0,61	0,99	2,92	5,34	1,50	-1,83
27	0,61	0,38	-0,39	0,41	0,24	0,40	<b>1,13</b>	0,40	1,36	-0,62	0,21	0,45
28	0,24	0,43	-0,31	0,43	<b>1,17</b>	0,40	1,01	0,45	1,59	-1,33	0,26	0,83
29	0,70	0,43	-0,71	0,39	0,39	0,47	<b>0,92</b>	0,43	1,41	-2,07	0,32	1,47
30	<b>1,40</b>	0,50	0,30	0,50	1,31	0,47	0,90	0,48	2,14	1,38	0,27	-0,65
32	<b>0,96</b>	0,36	0,38	0,41	0,86	0,40	0,47	0,39	1,42	1,62	0,24	-1,14
33	<b>2,50</b>	1,33	1,91	1,21	0,49	0,92	1,00	0,91	3,34	6,07	1,78	-1,82
35	<b>1,60</b>	0,47	0,20	0,47	-0,13	0,42	0,61	0,49	1,73	2,43	0,37	-1,41
37	0,62	0,30	0,05	0,31	0,22	0,28	<b>0,77</b>	0,28	1,01	-0,36	0,17	0,36
38	1,16	0,44	0,04	0,49	<b>1,60</b>	0,47	0,39	0,48	2,01	0,73	0,24	-0,36
40	<b>0,82</b>	0,29	0,11	0,34	0,06	0,28	0,68	0,29	1,07	-0,15	0,18	0,14
43	<b>0,84</b>	0,53	0,42	0,69	-0,35	0,78	<b>0,75</b>	0,84	1,25	-4,36	0,91	3,48
45	0,68	0,75	<b>1,11</b>	0,76	0,60	0,74	-0,21	0,74	1,45	4,44	0,72	-3,06
46	0,70	0,40	0,59	0,46	<b>1,40</b>	0,45	0,83	0,41	1,87	1,23	0,25	-0,66
47	1,37	1,28	<b>1,51</b>	1,01	0,21	0,86	0,14	1,09	2,05	5,40	1,38	-2,63
48	<b>1,89</b>	0,58	0,73	0,59	0,81	0,54	0,72	0,62	2,30	3,12	0,50	-1,36
52	0,82	0,38	0,63	0,44	<b>1,14</b>	0,43	0,24	0,39	1,56	1,82	0,27	-1,17
55	<b>1,09</b>	0,41	0,14	0,42	0,14	0,38	0,99	0,40	1,49	1,52	0,27	-1,02
56	<b>2,57</b>	1,28	1,03	1,35	0,66	1,14	0,84	1,16	2,97	6,78	2,46	-2,28
57	-1,98	0,97	<b>2,04</b>	1,07	-0,26	0,81	1,63	0,98	3,29	-3,05	0,91	0,93
59	-0,55	0,59	0,27	0,73	<b>2,00</b>	0,91	-0,18	0,65	2,10	3,94	0,99	-1,88
60	-1,30	1,00	0,20	1,19	0,22	1,00	<b>2,03</b>	1,53	2,43	5,74	2,17	-2,36
61	-0,86	0,41	0,62	0,49	0,79	0,42	0,90	0,43	1,60	0,20	0,20	-0,25
64	-0,51	0,82	0,47	1,10	<b>2,08</b>	1,01	-0,34	1,29	2,22	5,59	1,34	-2,52
65	-1,00	1,05	0,86	1,46	<b>2,58</b>	1,92	0,19	1,28	2,90	6,71	2,89	-2,31
66	0,73	0,74	0,06	1,13	<b>1,56</b>	1,05	0,31	0,72	1,75	5,02	1,28	-2,87
69	-0,25	0,39	0,19	0,40	-0,97	0,31	0,30	0,34	1,06	1,10	0,24	-1,04
70	-0,29	0,30	0,48	0,35	<b>0,63</b>	0,31	0,51	0,32	0,99	-0,57	0,20	0,58
71	-0,31	0,45	0,79	0,47	-0,15	0,44	<b>1,01</b>	0,45	1,33	2,28	0,32	-1,72
74	-1,14	0,56	1,18	0,65	-0,36	0,54	<b>1,29</b>	0,65	2,12	-1,77	0,36	0,84
75	-0,35	0,31	<b>0,79</b>	0,35	-0,33	0,31	0,43	0,35	1,02	0,53	0,17	-0,52

a 1 – Parâmetro de discriminação da dimensão 1; a 2 – Parâmetro de discriminação da dimensão 2; a 3 – Parâmetro de discriminação da dimensão 3; a 4 – Parâmetro de discriminação da dimensão 4; s.c. – erro padrão; MDISC – Parâmetro de discriminação multidimensional; d – Parâmetro de localização; MDIFF – Parâmetro de dificuldade multidimensional.

possui maior informação. Entretanto, como se trata de uma modelagem multidimensional compensatória, um item nem sempre é exclusivo de uma só dimensão, podendo contribuir para a estimação do traço latente do respondente (website) em várias dimensões.

De uma forma geral, a primeira dimensão agrupa itens com características voltadas para condução/orientação do usuário no website, também classificada, por alguns autores, como navegabilidade e usabilidade (González & Palacios, 2004; Nusair & Kandampully, 2008; Singh et al., 2016). Essa dimensão contém itens com carregamento secundário principalmente na dimensão 4, a qual é caracterizada por itens relacionados com apresentação da informação. Isso se deve ao fato de que, apesar de ser utilizada rotação ortogonal para apresentação dos resultados, existe algum tipo de relação entre as dimensões, e isso fica mais evidente nas dimensões 1 e 4, particularmente porque essas características são tratadas em alguns estudos (Tezza et al., 2011; Yang et al., 2004) como características altamente correlacionadas. Jun et al. (2004) classificam a organização da informação como facilidade de uso, que nesse contexto se pode entender como usabilidade. A dimensão 3 agrupa itens associados à interatividade.

No presente estudo, a correlação entre a dimensão 1 e a dimensão 4 foi de aproximadamente 0,4. O comportamento dessas duas dimensões reflete-se no carregamento dos itens 3, 19, 23, 35, 37 e 43.

Essa estrutura compensatória permite que o aumento de qualidade em uma dimensão possa compensar a redução em outra. No item 19, descrito como “as palavras aparentemente clicáveis são de fato clicáveis”, é nítido o conceito de navegação e orientação do usuário. Entretanto, esse item possui reflexos em outros conceitos como, por exemplo, o de apresentação da informação, ou seja, a apresentação de uma palavra ou frase transmite uma informação de funcionalidade que orienta e informa o usuário na interface.

A segunda dimensão evidenciou itens de acessibilidade, de forma semelhante à primeira dimensão da análise fatorial da Tabela 4, e confiabilidade/segurança. Os itens característicos dessa dimensão são os itens 12, 25, 45, 47, 57 e 75. O item 12 (existe informações de contato telefônico ou endereço) é uma característica importante para transmitir ao usuário confiabilidade e segurança de que, se ocorrer algum imprevisto, ele terá um endereço físico ou um telefone para recorrer (Nielsen & Tahir 2002). O item 25 (o site permite navegação em suas páginas em apenas uma janela, ou seja, não há abertura de novas janelas em meio a navegação) é um item ligado a confiabilidade e interação. Nielsen & Loranger (2006) afirmam que a maioria dos usuários não entende como manipular várias janelas e se concentra na que está mais à frente na tela. Se não puderem voltar, eles ficarão presos e

confusos. Nos itens 45 e 47, a questão relacionada com segurança fica mais evidente, uma vez que o primeiro item verifica se existe indicação de ambiente seguro no momento da efetivação da compra e o segundo verifica se o site possui outras formas de pagamento além do cartão de crédito. Os itens 57 e 75, conforme discutido na análise fatorial, na seção 4.3, estão ligados diretamente com a acessibilidade do sistema.

A dimensão 3 apresenta itens característicos de interatividade. Por exemplo, os itens 21 e 28 estão ligados diretamente ao mecanismo de busca do website, o qual representa uma ferramenta bastante comum que permite ao usuário interagir com o sistema efetuando sua busca por produtos e/ou informações. Os itens 64, 65 e 70 refletem o controle e poder de interação do usuário para com o sistema, possibilitando que o usuário possa pausar ou interromper algum conteúdo interativo ou de atualização automática.

A quarta dimensão, como discutido anteriormente, possui correlação com a primeira dimensão. Entretanto, existem itens característicos que a caracterizam como sendo apresentação da informação, como, por exemplo, os itens 27, 37, 60 e 74.

Na grande maioria das aplicações da teoria da resposta ao item, seja ela unidimensional ou multidimensional, os parâmetros de discriminação do modelo são positivos. Isso ocorre principalmente por questões relacionadas ao próprio construto, ou seja, assume-se que um valor positivo de parâmetro de discriminação resulta em uma monotonicidade crescente, ou seja, com o aumento da proficiência do indivíduo aumenta a probabilidade de resposta positiva ao item. Esse comportamento é bastante comum nas áreas educacional e psicológica, nas quais a TRI é mais difundida. Entretanto, é possível se ter valores do parâmetro de discriminação (parâmetro  $a$ ) negativos, tal situação mantém o pressuposto da monotonicidade, porém decrescente.

A Figura 1 mostra a estrutura criada pelo modelo compensatório multidimensional da teoria da resposta ao item. Para construção desse diagrama assumiu-se o carregamento em cada dimensão com base no sentido da carga de discriminação e na sua intensidade, assumindo-se aqui como sendo relevantes cargas maiores que 0,70. Segundo Hair et al. (2009), em uma análise fatorial, cargas fatoriais na faixa de +0,30 a +0,40 são consideradas como tendo nível mínimo para interpretação de estrutura. Cargas de +0,50 ou maiores são tidas como praticamente significantes e maiores de +0,70 são consideradas indicativas de estrutura bem definida. Dessa forma, optou-se por considerar como relevantes cargas maiores que 0,70.

A Figura 1 demonstra parte da complexidade que uma modelagem multidimensional compensatória envolve. As linhas cheias representam a interação de maior carga discriminatória e as pontilhadas, as

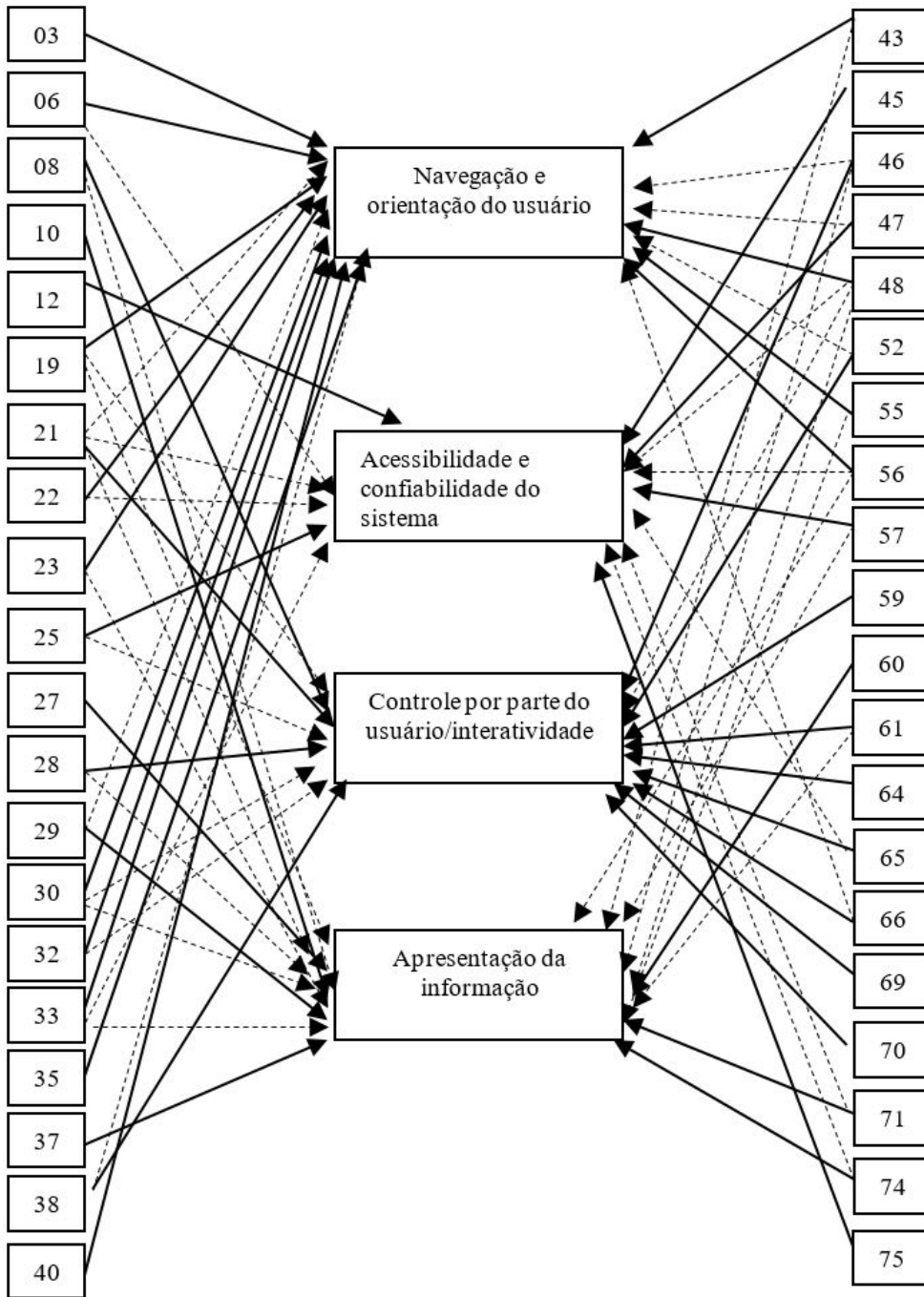


Figura 1. Representação gráfica da relação de cada um dos 40 itens nas 4 dimensões caracterizadas.

interações secundárias com carga discriminatória no mesmo sentido da principal e com intensidade maior que 0,70. Por exemplo, o item 47 (O site possui outras formas de pagamento além do cartão de crédito) possui maior discriminação na dimensão acessibilidade e confiabilidade com  $a = 1,57$  e interação secundária na dimensão navegação e orientação com  $a = 1,37$ . Do ponto de vista prático, representa que a característica de possuir outras formas de pagamento além do cartão de crédito

contribui do ponto de vista de qualidade em um website em termos de confiabilidade, ou seja, o website possibilita e permite ao usuário escolher a forma de pagamento que lhe pareça mais conveniente e segura, sendo, portanto, uma característica comum às duas dimensões, como tantas outras características apresentada na Figura 1.

O Quadro 2 mostra a distribuição dos itens quanto às dimensões caracterizadas, considerando-se as maiores cargas.

## 5.4 Interpretação da escala gerada

Com base na estimação dos parâmetros de discriminação multidimensional é possível classificar cada item conforme seu poder de discriminação ou seu poder informativo no construto, bem como, com base no parâmetro de dificuldade, classificar cada item conforme sua dificuldade.

Na Tabela 5 verifica-se que os itens 43, 29, 57, 74 e 28 são os itens que exigem mais de um website, ou seja, são os cinco itens mais difíceis do construto. Do ponto de vista prático, isso faz sentido. Por exemplo, o item 43 é relativo a multimídia para apresentação de produtos. Essa é uma característica pouco comum na maioria dos websites, encontrada apenas em websites mais maduros do ponto de vista de qualidade da apresentação da informação. O mesmo acontece com os itens 28 e 29, os quais tratam do aperfeiçoamento da ferramenta de busca.

Quanto ao ranque de discriminação, na Tabela 5 verifica-se que os cinco itens mais discriminantes foram os itens 33, 57, 21, 56 e 25.

A Tabela 6 mostra a estimação da habilidade dos quatro primeiros websites analisados, na escala normal  $N(0;1)$ , ou seja, média zero e variância um,

caracterizando estatisticamente, portanto, uma escala padronizada. Verifica-se que o primeiro website possui um maior domínio dos itens relacionados à apresentação da informação, precisando, assim, desenvolver melhor sua qualidade, principalmente no controle por parte do usuário/interatividade, a qual apresentou domínio abaixo da média. No website número 2 observa-se que existe um bom domínio da acessibilidade/confiabilidade do sistema, havendo, entretanto, necessidade de melhoramento no que se refere aos requisitos de controle por parte do usuário/interatividade e apresentação da informação.

Entretanto, em um modelo multidimensional, a análise pontual das estimativas das proficiências dos respondentes, nesse caso o grau de qualidade de um website pode remeter a uma avaliação simplista do modelo, uma vez que a estrutura multidimensional agrega mais informação e complexidade que um modelo unidimensional. Sendo assim, cada dimensão pode ser avaliada em particular, considerando-se cada item associado ao grau de qualidade estimado de cada website, como pode ser visualizado na Tabela 7. Essa tabela traz a probabilidade de acerto (resposta positiva) de cada item, com base nas estimativas do

**Quadro 2.** Classificação dos itens conforme a dimensão, segundo o modelo multidimensional da teoria da resposta ao item.

Dimensão	Itens
Condução do usuário/navegabilidade	03, 06, 19, 22, 23, 30, 32, 33, 35, 40, 43, 48, 55, 56
Acessibilidade/confiabilidade do sistema	12, 25, 45, 47, 57, 75
Controle por parte do usuário/interatividade	08*, 21, 28, 38, 46, 52, 59, 64, 65, 66, 69*, 70
Apresentação da informação	10, 27, 29, 37, 60, 61, 71, 74

\*Itens com carregamento contrário à maioria dos itens da dimensão.

**Tabela 6.** Estimação do grau de qualidade multidimensional dos quatro primeiros websites da amostra com base no modelo multidimensional de dois parâmetros compensatório.

Website	Orientação do usuário/navegabilidade	Acessibilidade/confiabilidade do sistema	Controle por parte do usuário/interatividade	Apresentação da informação
001	0,105	0,130	-0,864	0,677
002	0,367	0,524	-0,753	0,040
003	-0,887	-0,328	-0,557	0,525
004	0,782	0,324	-0,879	0,314

**Tabela 7.** Estimativa do grau de qualidade multidimensional dos quatro primeiros sites da amostra com base no modelo multidimensional de dois parâmetros compensatório.

Website	Orientação do usuário/navegabilidade	Acessibilidade/confiabilidade do sistema	Controle por parte do usuário/interatividade	Apresentação da informação
001	0,105	0,130	-0,864	0,677
002	0,367	0,524	-0,753	0,040
003	-0,887	-0,328	-0,557	0,525
004	0,782	0,324	-0,879	0,314

grau de qualidade dos quatro primeiros websites, conforme Tabela 6.

Do ponto de vista prático, a análise apresentada na Tabela 7 demonstra, com base no grau de qualidade do website estimado pelo modelo multidimensional compensatório de quatro dimensões, os itens que são teoricamente dominados pelo website. Considerando o conceito de itens âncoras definidos por Beaton & Allen (1992) e Andrade et al. (2000), no qual um dos critérios para caracterizar um item em determinado grau da escala é a probabilidade maior do que 0,650 naquele ponto da escala, verifica-se que o website 1 domina a maioria dos itens, exceto os itens 27, 28, 29, 37, 38, 40, 43, 57, 61, 70 e 74, os quais apresentam probabilidade de resposta positiva inferior a 0,650. Observa-se que, apesar de o website 1 apresentar maior grau de qualidade na dimensão 4, ele tem, assim como os outros três websites mostrados na Tabela 7, probabilidade abaixo de 0,650 ( $p = 0,511$ ) de responder positivamente ao item 37, o qual trata, segundo o modelo, de uma característica associada a apresentação da informação. Isso se dá particularmente por se tratar de um item com parâmetro de dificuldade acima da média (0,360). Tal análise pode ser seguida aos demais itens.

A análise de cada website associado com a probabilidade de resposta positiva a cada item amplia a análise prática da modelagem multidimensional, pois particulariza cada item e demonstra onde, pontualmente, melhorias precisam ser implementadas. Já a análise geral do grau de qualidade apresentado na Tabela 6 dá um indicativo relativo à dimensão ou ao conceito nos quais o website precisa melhorar ou deve explorar mais.

## 6 Considerações finais

O trabalho desenvolveu um modelo multidimensional utilizando a teoria da resposta ao item para mensurar qualidade em websites comerciais. Para criação do modelo utilizou-se como base o modelo logístico de dois parâmetros multidimensional com quatro dimensões e um conjunto de 40 itens. As quatro dimensões mostraram-se associadas a conceitos de navegabilidade ou condução/orientação do usuário, acessibilidade e confiabilidade do sistema, interatividade e apresentação da informação. Essas dimensões encontradas no presente trabalho remetem às dimensões mais citadas na literatura e relacionadas diretamente com a definição de qualidade em websites, que é um conjunto de características técnicas e não técnicas de um sistema web que permitem ao usuário realizar seus objetivos em um website de forma acessível, eficiente e agradável. Como características técnicas, entende-se usabilidade/navegabilidade, apresentação da informação, acessibilidade e interatividade do sistema. Como características não técnicas, *design*,

estética, apelo visual/comercial, confiabilidade, hedonismo, empatia.

Do ponto de vista de implicações teóricas ao campo de qualidade em websites, o presente trabalho desenvolveu um modelo para avaliar requisitos de qualidade, discutiu questões relativas à dimensionalidade e sugeriu um modelo multidimensional capaz de avaliar individualmente cada item e cada website, nas respectivas dimensões.

Do ponto de vista prático, o trabalho abre outras possibilidades de mensuração da qualidade em websites comerciais, com a finalidade de identificar melhorias possíveis da interface e, possivelmente, o desenvolvimento de ferramentas automatizadas para diagnósticos úteis para melhoria de projetos de website.

## Referências

- AChecker. (2012). Web accessibility checker. Recuperado em 25 de maio de 2012, de <https://achecker.ca/checker/index.php>
- Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, 29(1), 67-91. <http://dx.doi.org/10.1111/j.1745-3984.1992.tb00368.x>.
- Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, 7(4), 255-278. [http://dx.doi.org/10.1207/s15324818ame0704\\_1](http://dx.doi.org/10.1207/s15324818ame0704_1).
- Adams, R. J., Wilson, M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21(1), 1-23. <http://dx.doi.org/10.1177/0146621697211001>.
- Agarwal, R., & Venkatesh, V. (2002). Assessing a firm's web presence: a heuristic evaluation procedure for the measurement of usability. *Information Systems Research*, 13(2), 168-186. <http://dx.doi.org/10.1287/isre.13.2.168.84>.
- Aladwani, A. M., & Palvia, P. C. (2002). Developing and validating an instrument for measuring user-perceived web quality. *Information & Management*, 39(6), 467-476. [http://dx.doi.org/10.1016/S0378-7206\(01\)00113-6](http://dx.doi.org/10.1016/S0378-7206(01)00113-6).
- Al-Khalifa, H. S. (2010). The accessibility of Saudi Arabia government Web sites: an exploratory study. *Universal Access in the Information Society*, 10(4). <http://dx.doi.org/10.1007/s10209-016-0495-7>.
- Andrade, D. F., Tavares, H. R., & Valle, R. C. (2000). *Teoria de resposta ao item: conceitos e aplicações*. São Paulo: ABE.
- Babbie, E. (2005). *The basics of social research*. Belmont: Wadsworth Publishing.
- Baker, D. L. (2009). Advancing E-Government performance in the United States through enhanced usability benchmarks.

- Government Information Quarterly*, 26(1), 82-88. <http://dx.doi.org/10.1016/j.giq.2008.01.004>.
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24(6), 574-594. <http://dx.doi.org/10.1080/10447310802205776>.
- Barnes, S. J., & Vidgen, R. T. (2000). WebQual: an exploration of web site quality. In *Proceedings of the 8th European Conference on Information Systems* (Vol. 1, pp. 298-305). Vienna: Wirtschaftsuniversität Wien.
- Bartolucci, F., Montanari, G. E., & Pandolfi, S. (2012). Dimensionality of the latent structure and item selection via latent class multidimensional IRT models. *Psychometrika*, 77(4), 782-802. <http://dx.doi.org/10.1007/s11336-012-9278-0>.
- Beaton, A. E., & Allen, N. L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics*, 17(2), 191-204.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's Ability. In F. M. Lord, & M. R. Novick. *Statistical theories of mental test scores* (Vol. 38, pp. 123-140). Boston: Addison-Wesley.
- Braz, C., Seffah, A., & M'Raihi, D. (2007). Designing a trade-off between usability and security: a metrics based-model. *Lecture Notes in Computer Science*, 4663, 114-126. [http://dx.doi.org/10.1007/978-3-540-74800-7\\_9](http://dx.doi.org/10.1007/978-3-540-74800-7_9).
- Cai, L. (2012). *FlexMIRTTM version 1.86: a numerical engine for multilevel item factor analysis and test scoring*. Seattle: Vector Psychometric Group.
- Camilli, G. (1992). A conceptual analysis of differential item functioning in terms of a multidimensional item response model. *Applied Psychological Measurement*, 16(2), 129-147. <http://dx.doi.org/10.1177/014662169201600203>.
- Chalmers, R. P. (2012). Mirt: a multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29. <http://dx.doi.org/10.18637/jss.v048.i06>.
- Cristobal, E., Flavian, C., & Guinaliu, M. (2007). Perceived e-service quality (PeSQ): measurement validation and effects on consumer satisfaction and web site loyalty. *Managing Service Quality*, 17(3), 317-340. <http://dx.doi.org/10.1108/09604520710744326>.
- Day, A. (1997). A model for monitoring Web site effectiveness. *Internet Research: Electronic Networking Applications and Policy*, 7(2), 1-9. <http://dx.doi.org/10.1108/10662249710165244>.
- DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of Management Information Systems*, 19(4), 9-30. <http://dx.doi.org/10.1080/07421222.2003.11045748>.
- Deng, L., & Poole, M. S. (2012). Aesthetic design of e-commerce web pages – webpage complexity, order and preference. *Electronic Commerce Research and Applications*, 11(4), 420-440. <http://dx.doi.org/10.1016/j.elerap.2012.06.004>.
- Ding, D. X., Hu, P. J. H., & Sheng, O. R. L. (2011). e-SELFQUAL: a scale for measuring online self-service quality. *Journal of Business Research*, 64(1), 508-515. <http://dx.doi.org/10.1016/j.jbusres.2010.04.007>.
- Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, 56(3), 495-515. <http://dx.doi.org/10.1007/BF02294487>.
- Embretson, S., & Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates, Inc. Publishers.
- Éthier, J., Hadaya, P., Talbot, J., & Cadieux, J. (2008). Interface design and emotions experienced on B2C Web sites: empirical testing of a research model. *Computers in Human Behavior*, 24(2), 2771-2791. <http://dx.doi.org/10.1016/j.chb.2008.04.004>.
- Fang, X., & Holsapple, C. W. (2007). An empirical study of web site navigation structures' impacts on web site usability. *Decision Support Systems*, 43(2), 476-491. <http://dx.doi.org/10.1016/j.dss.2006.11.004>.
- Fassnacht, M., & Koesel, I. (2006). Quality of electronic services: conceptualizing and testing a hierarchical model. *Journal of Service Research*, 9(1), 19-37. <http://dx.doi.org/10.1177/1094670506289531>.
- Flavián, C., & Guinaliú, M. (2006). Consumer trust, perceived security and privacy policy: three basic elements of loyalty to a web site. *Industrial Management & Data Systems*, 106(5), 601-620. <http://dx.doi.org/10.1108/02635570610666403>.
- Flavián, C., Guinaliú, M., & Gurrea, R. (2006). The influence of familiarity and usability on loyalty to online journalistic services: the role of user experience. *Journal of Retailing and Consumer Services*, 13(5), 363-375. <http://dx.doi.org/10.1016/j.jretconser.2005.11.003>.
- Fleck, M. P. A., & Bourdel, M. C. (1998). Método de simulação e escolha de fatores na análise dos principais componentes. *Revista de Saude Publica*, 32(3), 267-272. <http://dx.doi.org/10.1590/S0034-89101998000300010>. PMID:9778862.
- Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: a critical review and analysis. *Personnel Psychology*, 39(2), 292-314. <http://dx.doi.org/10.1111/j.1744-6570.1986.tb00583.x>.
- Galletta, D. F., Henry, R., McCoy, S., & Polak, P. (2004). Web site delays: how tolerant are users? *Journal of the Association for Information Systems*, 5(1), 1-28.
- Gata, W., & Gilang, O. (2017). Analysis of information system quality of service on bsi academy's environment using webqual methods, importance performance analysis and fishbone. *Journal of Theoretical and Applied Information Technology*, 95(2), 229.



- Gefen, D. (2002). Consumer loyalty in e-commerce. *Journal of the Association for Information Systems*, 3(1), 27-51. <http://dx.doi.org/10.17705/1jais.00022>.
- Glas, C. A. W. (1992). A rasch model with a multivariate distribution of ability. In M. Wilson (Ed.), *Objective measurement: theory into practice* (pp. 236-258). Norwood: Ablex.
- Goldberg, C. B., & Allen, D. G. (2008). Black and white and read all over: Race differences in reactions to recruitment Web sites. *Human Resource Management*, 47(2), 217-236. <http://dx.doi.org/10.1002/hrm.20209>.
- González, F. J. M., & Palacios, T. M. B. (2004). Quantitative evaluation of commercial websites: an empirical study of Spanish firms. *International Journal of Information Management*, 24(4), 313-328. <https://doi.org/10.1016/j.ijinfomgt.2004.04.009>.
- Gummerus, J., Liljander, V., Pura, M., & van Riel, A. (2004). Customer loyalty to content-based web sites: the case of an online health-care service. *Journal of Services Marketing*, 18(3), 175-186. <http://dx.doi.org/10.1108/08876040410536486>.
- Haga, W. J., & Zviran, M. (1994). Information systems effectiveness: research designs for causal inference. *Journal of Information Systems*, 4(2), 141-166. <http://dx.doi.org/10.1111/j.1365-2575.1994.tb00048.x>.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Análise multivariada de dados* (6. ed.). Porto Alegre: Bookman.
- Hartig, J., & Höhler, J. (2008). Representation of competencies in multidimensional IRT models with within-item and between-item multidimensionality. *The Journal of Psychology*, 216(2), 88-100.
- Hasan, B. (2016). Perceived irritation in online shopping: the impact of website design characteristics. *Computers in Human Behavior*, 54, 224-230. <http://dx.doi.org/10.1016/j.chb.2015.07.056>.
- Ibrahim, E. E., Joseph, M., & Ibeh, K. I. N. (2006). Customers' perception of electronic service delivery in the UK retail banking sector. *International Journal of Bank Marketing*, 24(7), 475-493. <http://dx.doi.org/10.1108/02652320610712094>.
- International Organisation for Standardization – ISO. (1992). *ISO9126: Information Technology — Software Product Evaluation — Quality Characteristics and Guidelines for Their Use*. Geneva: ISO.
- Johnson, R., & Wichern, D. (2007). *Applied multivariate analysis* (6th. ed.). New Jersey: Prentice Hall.
- Jun, M., Yang, Z., & Kim, D. S. (2004). Customers' perceptions of online retailing service quality and their satisfaction. *International Journal of Quality & Reliability Management*, 21(8), 817-840. <http://dx.doi.org/10.1108/02656710410551728>.
- Kaplan, D., Krishnan, R., Padman, R., & Peters, J. (1998). KRISHNAN, R. PADMAN, R. PETERS, J. Assessing data quality in accounting information systems. *Communications of the ACM*, 41(2), 72-78. <http://dx.doi.org/10.1145/269012.269024>.
- Kettinger, W. J., & Lee, C. C. (1994). Perceived service quality and user satisfaction with the information services function. *Decision Sciences*, 25(5-6), 737-766. <http://dx.doi.org/10.1111/j.1540-5915.1994.tb01868.x>.
- Kim, C., Galliers, R. D., Shin, N., Ryoo, J., & Kim, J. (2012). Factors influencing Internet shopping value and customer repurchase intention. *Electronic Commerce Research and Applications*, 11(4), 374-387. <http://dx.doi.org/10.1016/j.elerap.2012.04.002>.
- Kim, S., & Stoel, L. (2004). Dimensional hierarchy of retail website quality. *Information & Management*, 41(5), 619-633. <http://dx.doi.org/10.1016/j.im.2003.07.002>.
- King, W. R., & Epstein, B. J. (1983). Assessing information system value: an experimental study. *Decision Sciences*, 14(1), 34-45. <http://dx.doi.org/10.1111/j.1540-5915.1983.tb00167.x>.
- Kitchenham, B., & Pflieger, S. L. (1996). Software quality: the elusive target. *IEEE Software*, 13(1), 12-21. <http://dx.doi.org/10.1109/52.476281>.
- Kline, P. (2000). *The handbook of psychological testing*. London: Routledge.
- Kulviwat, S., Bruner, G. C., 2nd., Kumar, A., Nasco, S. A., & Clark, T. (2007). Toward a unified theory of consumer acceptance of technology. *Psychology and Marketing*, 24(12), 1067-1092. <http://dx.doi.org/10.1002/mar.20196>.
- Lindgaard, G. (2007). Aesthetics, visual appeal, usability, and user satisfaction: What do the user's eyes tell the user's brain? *Australian Journal of Emerging Technologies and Society*, 5(1), 1-14.
- Lindgaard, G., Gary, J. F., Cathy, D., & Brown, J. (2006). Attention Web Designers: You Have 50 Milliseconds to Make a Good First Impression! *Behaviour & Information Technology*, 25(2), 115-126. <http://dx.doi.org/10.1080/01449290500330448>.
- Lindroos, K. (1997). Use quality and the World Wide Web. *Information and Software Technology*, 39(12), 827-836. [http://dx.doi.org/10.1016/S0950-5849\(97\)00041-4](http://dx.doi.org/10.1016/S0950-5849(97)00041-4).
- Loiacono, E., Watson, R., & Goodhue, D. (2002). WEBQUAL: a measure of web site quality. In K. Evans, & L. Scheer (Eds.), *Marketing Educators' Conference: Marketing Theory and Applications* (Vol. 13, pp. 432-437). Amsterdam: Elsevier Science B.V.
- Lord, F. M. (1952). *A theory of test scores* (Psychometric Monograph, No. 7). Madison: Psychometric Society.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale: Erlbaum.
- McDonald, R. P. (1981). The dimensionality of test and items. *The British Journal of Mathematical and Statistical Psychology*, 33(1), 161-183.

- Nelson, K. G. (1996). Global information systems quality: key issues and challenges. *Journal of Global Information Management*, 4(4), 4-14. <http://dx.doi.org/10.4018/jgim.1996100101>.
- Nielsen, J., & Loranger, H. (2006). *Prioritizing web usability*. California: New Riders.
- Nielsen, J., & Tahir, M. 2002. *Homepage usability: 50 sites deconstructed*. Indianapolis: New Riders Publishing.
- Nusair, K., & Kandampully, J. (2008). The antecedents of customer satisfaction with *online* travel services: a conceptual model. *European Business Review*, 20(1), 4-19. <http://dx.doi.org/10.1108/09555340810843663>.
- Oshima, T. C., & Miller, M. D. (1992). Multidimensionality and item bias in item response theory. *Applied Psychological Measurement*, 16(3), 237-248. <http://dx.doi.org/10.1177/014662169201600304>.
- Palmer, J. W. (2002). Web site usability, design, and performance metrics. *Information Systems Research*, 13(2), 151-167. <http://dx.doi.org/10.1287/isre.13.2.151.88>.
- Pandir, M., & Knight, J. (2006). Homepage aesthetics: the search for preference factors and the challenges of subjectivity. *Interacting with Computers*, 18(6), 1351-1370. <http://dx.doi.org/10.1016/j.intcom.2006.03.007>.
- Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). E-S-Qual: a multiple-item scale for assessing electronic service quality. *Journal of Service Research*, 7(3), 213-233. <http://dx.doi.org/10.1177/1094670504271156>.
- Pinterits, A., Treiblmaier, H., & Pollach, I. (2006). Environmental websites: an empirical investigation of functionality and accessibility. *International Journal of Technology Policy and Management*, 6(1), 103-119. <http://dx.doi.org/10.1504/IJTPM.2006.010075>.
- Podsakoff, P. M., Ahearne, M., & Mackenzie, S. B. (1997). Organizational citizenship behaviors and the quantity and quality of work group performance. *The Journal of Applied Psychology*, 82(2), 262-270. <http://dx.doi.org/10.1037/0021-9010.82.2.262>. PMID:9109284.
- R Core Team. (2012). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Recuperado em 25 de maio de 2012, de <http://www.R-project.org/>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Rauch, D., & Hartig, J. (2010). Multiple-choice versus open-ended response formats of reading test items: a two-dimensional IRT analysis. *Psychological Test and Assessment Modeling*, 52(4), 354-379.
- Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, 9(4), 401-412. <http://dx.doi.org/10.1177/014662168500900409>.
- Reckase, M. D. (1997). A linear logistic multidimensional model for dichotomous items response data. In W. J. van der Linden & W. J. Hambleton (Ed.), *Handbook of modern item response theory* (pp. 271-286). New York: Springer.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer. <http://dx.doi.org/10.1007/978-0-387-89976-3>.
- Reckase, M. D., & McKinley, R. L. (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement*, 15(4), 361-373. <http://dx.doi.org/10.1177/014662169101500407>.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114(3), 552-566. <http://dx.doi.org/10.1037/0033-2909.114.3.552>. PMID:8272470.
- Revelle, W. (2012). *Package 'psych': procedures for psychological, psychometric, and personality research. Version 1.2. 12*. Recuperado em 25 de maio de 2012, de <http://personality-project.org/r/psych.manual.pdf>
- Rust, R. T. (2001). The rise of E-Service. *Journal of Service Research*, 3(4), 283-284. <http://dx.doi.org/10.1177/109467050134001>.
- Schmettow, M., & Vietze, W. (2008). Introducing item response theory for measuring usability inspection processes. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*. New York: ACM. <http://dx.doi.org/10.1145/1357054.1357196>.
- Schneidewind, N. F. (1992). Methodology for validating software metrics. *IEEE Transactions on Software Engineering*, 18(5), 410-422. <http://dx.doi.org/10.1109/32.135774>.
- Seffah, A., Donyaee, M., Kline, R. B., & Padda, H. K. (2006). Usability measurement and metrics: a consolidated model. *Software Quality Control*, 14(2), 159-178. <http://dx.doi.org/10.1007/s11219-006-7600-8>.
- Serkan, A., Eda, A., & Safak, A. (2010). Re-assessment of E-S-Qual and E-RecS-Qual in a pure service setting. *Journal of Business Research*, 63(3), 232-240. <http://dx.doi.org/10.1016/j.jbusres.2009.02.018>.
- Silva, R. G. S., Frio, R. S., Sampaio, C. H., & Broilo, P. L. (2015). Qualidade percebida do autosserviço online: validação da escala e-SELFQUAL no contexto brasileiro. *Revista Brasileira de Marketing*, 14(2), 138-154.
- Singh, T., Malik, S., & Sarkar, D. (2016). E-commerce website quality assessment based on usability. In *International Conference on Computing, Communication and Automation (ICCCA)* (pp. 101-105). USA: IEEE.
- Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artificial factors. *Journal of Management*, 23(5), 659-677. <http://dx.doi.org/10.1177/014920639702300503>.

- Stepchenkova, S., Tang, L., Jang, S. S., Kirilenko, A. P., & Morrison, A. M. (2010). Benchmarking CVB website performance: Spatial and structural patterns. *Tourism Management*, 31(5), 611-620. <http://dx.doi.org/10.1016/j.tourman.2009.06.015>.
- Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., & Mani, M. (2015). Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR mHealth and uHealth*, 3(1), e27. <http://dx.doi.org/10.2196/mhealth.3422>. PMID:25760773.
- Tavares, H. R., Andrade, D. F., & Pereira, C. A. B. (2004). Detection of determinant genes and diagnostic via item response theory. *Genetics and Molecular Biology*, 27(4), 679-685. <http://dx.doi.org/10.1590/S1415-47572004000400033>.
- Taylor, M. J., & England, D. (2006). Internet marketing: web site navigational design issues. *Marketing Intelligence & Planning*, 24(1), 77-85. <http://dx.doi.org/10.1108/02634500610641570>.
- Tezza, R., Bornia, A. C., & Andrade, D. F. (2011). Measuring web usability using item response theory: principles, features and opportunities. *Interacting with Computers*, 23(2), 167-175. <http://dx.doi.org/10.1016/j.intcom.2011.02.004>.
- Tezza, R., Bornia, A. C., Spenassato, D., & Trierweiler, A. C. (2016). Empirical comparison of the multidimensional models of item response theory in e-commerce. *Pesquisa Operacional*, 36(3), 503-532. <http://dx.doi.org/10.1590/0101-7438.2016.036.03.0503>.
- van der Merwe, R., & Bekker, J. (2003). A framework and methodology for evaluating e-commerce Web sites. *Internet Research Electronic Networking Applications and Policy*, 13(5), 330-341. <http://dx.doi.org/10.1108/10662240310501612>.
- W3C. (2008). *Web content accessibility guidelines (WCAG) 2.0. W3C Recommendation*. Recuperado em 25 de maio de 2012, de <http://www.w3.org/TR/WCAG20/>
- Wang, R., Storey, V., & Firth, A. (1995). A framework for data quality research. *IEEE Transactions on Knowledge and Data Engineering*, 7(4), 623-640. <http://dx.doi.org/10.1109/69.404034>.
- Xie, M., Wang, H., & Goh, T. N. (1998). Quality dimensions of Internet search engines. *Journal of Information Science*, 24(5), 365-372. <http://dx.doi.org/10.1177/016555159802400509>.
- Yang, Z., Cai, S., Zhou, Z., & Zhou, N. (2005). Development and validation of an instrument to measure user perceived service quality of information presenting Web portals. *Information & Management*, 42(4), 575-589.
- Yang, Z., Jun, M., & Peterson, R. T. (2004). Measuring customer perceived online service quality: scale development and managerial implications. *International Journal of Operations & Production Management*, 21(11), 1149-1174. <http://dx.doi.org/10.1108/01443570410563278>.
- Zeithaml, V. A., Parasuraman, A., & Malhotra, A. (2000). *E-service quality: definition, dimensions and conceptual model* (Working Paper). Cambridge: Marketing Science Institute.