

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO**

Rafael Tezza

**Proposta de um construto para medir usabilidade em *sites* de
e-commerce utilizando a Teoria da Resposta ao Item**

Dissertação de mestrado

**Florianópolis
2009**

RAFAEL TEZZA

**Proposta de um construto para medir usabilidade em *sites* de
e-commerce utilizando a Teoria da Resposta ao Item**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina como requisito parcial para obtenção de grau de Mestre em Engenharia de Produção.

Orientador: Prof. Antonio Cezar Bornia, Dr.

Florianópolis
2009

Rafael Tezza

**Proposta de um construto para medir usabilidade em *sites* de
e-commerce utilizando a Teoria da Resposta ao Item**

Esta dissertação foi julgada e aprovada para a obtenção do título de Mestre em Engenharia de Produção no Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina.

Florianópolis, 17 de fevereiro de 2009.

Prof. Antônio Sérgio Coelho, Dr.
Coordenador do Curso

Prof. Antonio Cezar Bornia, Dr.
DEPS/UFSC
Orientador

BANCA EXAMINADORA

Prof. Dalton Francisco de Andrade, PhD.
INE/UFSC

Prof^a. Leila Amaral Gontijo, Dra.
DEPS/UFSC

Prof. Adriano Ferreti Borgatto, Dr.
INE/UFSC

Aos que vêm na simplicidade e na objetividade
uma oportunidade de ir além.

AGRADECIMENTOS

Ao meu orientador, prof. Antonio Cezar Bornia pelo incentivo, cobrança e ensinamentos.
Ao prof. Dalton Francisco de Andrade por compartilhar seus conhecimentos e sua paixão pela Teoria da Resposta ao Item.
A Rafael Marques de Albuquerque, pelo companheirismo, pelas valiosas sugestões e revisões e por saber se fazer presente sempre.
Aos colegas do curso, e aos amigos Ivan Henrique Vey, Mariele Katherine Jungles.
Ao prof. Adriano Ferreti Borgatto, a prof^a Leila Amaral Gontijo e a todos os professores e funcionários do Programa de Pós-Graduação em Engenharia de Produção da UFSC.
A minha mãe e a minha avó, que foram as primeiras a me incentivar e me apoiar neste caminho que escolhi.
Ao CNPq pelo financiamento neste último ano do curso e a todos que ajudaram ou torceram para que este trabalho se tornasse real.

*Há um tempo em que é preciso abandonar as roupas usadas ...
Que já têm a forma do nosso corpo ...
E esquecer os nossos caminhos que nos levam sempre aos
mesmos lugares ...*

*É o tempo da travessia ...
E se não ousarmos fazê-la ...
Teremos ficado ... para sempre ...
À margem de nós mesmos...*

Fernando Pessoa

RESUMO

TEZZA, Rafael. **Proposta de um construto para medir usabilidade em *site* de *e-commerce* utilizando a Teoria da Resposta ao Item**. 2009. 139 f. Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Centro Tecnológico, Florianópolis, 2009.

O presente trabalho tem como objetivo avaliar a usabilidade de *sites* de comércio eletrônico, através da elaboração de um construto baseado em um questionário validado pela Teoria da Resposta ao Item (TRI). Considerando que um levantamento da literatura demonstrou que a maioria das abordagens sobre usabilidade em comércio eletrônico trata de questões muito mais subjetivas e baseadas em questionários de satisfação de clientes, o presente trabalho se propõe a elaboração de um construto que reduza este problema de subjetividade. Atualmente com o crescimento da economia digital, a questão de usabilidade, principalmente nos *sites* de comércio eletrônico, tem figurado como uma característica capaz de gerar potencial competitivo às empresas do setor. A pesquisa bibliográfica no campo de usabilidade e comércio eletrônico serviu de base para a elaboração do questionário, e a TRI identificou a relevância de cada item, em uma amostragem de cerca de 361 *sites* de comércio eletrônico. Como resultado, questões de usabilidade até então puramente teóricas ou baseadas em testes com usuários foram discutidas sob o ponto de vista matemático da TRI, e uma escala padronizada para medir usabilidade em *sites* de comércio eletrônico foi desenvolvida. A pesquisa apresentou algumas limitações, tais como a priorização de questões objetivas e a avaliação concentrada em apenas um usuário – o autor. Entretanto, tanto os resultados como as limitações representam oportunidade futuras de pesquisa sob este ponto de vista até agora inexplorado na avaliação de *sites* de comércio eletrônico.

Palavras-chave: comércio eletrônico, usabilidade, economia digital, teoria da resposta ao item.

ABSTRACT

TEZZA, Rafael. **Proposta de um construto para medir usabilidade em *site* de *e-commerce* utilizando a Teoria da Resposta ao Item**. 2009. 139 f. Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Centro Tecnológico, Florianópolis, 2009.

This work seeks to evaluate the usability of e-commerce sites through the elaboration of a questionnaire construct using the Item Response Theory (IRT). Taking into account that the Review of Literature has demonstrated that most of the approaches on the usability of e-commerce deal with much more subjective issues and are based on clients' satisfaction questionnaires, this work suggests the elaboration of a construct that reduces the subjectivity issue. At present, due to the digital economy growth, the usability issue, mainly at e-commerce sites, has figured as a characteristic capable of generating competitive power to organizations in such a sector. The bibliographical references in the field of usability and e-commerce have served as the basis for the elaboration of the questionnaire, and the IRT has identified the relevance of each item in a sample of about 361 e-commerce sites. As a result, usability matters so far considered as either theoretical matters or based on tests with users were discussed under the mathematical IRT point of view, and a patterned scale to measure the usability in e-commerce sites was developed. The investigation had some limitations, such as prioritizing objective issues as well as the evaluation focusing on only one user – the author. However, both the results and the limitations represent future opportunity for research following such a point of view so far unexplored on the evaluation of e-commerce sites.

Keywords: e-commerce, usability, digital economy, response to item theory.

LISTA DE FIGURAS

Figura 2.1 - Estrutura da usabilidade.....	19
Figura 2.2 – Tipos de Comércio Eletrônico	41
Figura 2.3 - Principais estágios de um survey	47
Figura 2.4 – Curva Característica do Item (CCI)	52
Figura 4.1 - Curva Característica do item 40	83
Figura 4.2 – Curva Característica do item 26	83
Figura 4.3 – Curva Característica e curva de informação do item 42.....	84
Figura 4.4 – Curva de informação total.....	85
Figura 4.5 - Distribuição do grau de usabilidade dos <i>sites</i> utilizados.....	86
Figura 4.6 - Frequência de <i>sites</i> em cada nível da escala do grau de usabilidade na métrica (50,10).....	102

LISTA DE QUADROS

Quadro 2.1 – Critérios ergonômicos considerados pelos autores segundo as características base	30
Quadro 2.2 - Descrição, vantagens e desvantagens das técnicas de avaliação de ergonomia de interface e usabilidade estudadas	39
Quadro 3.1 - Os itens e suas descrições.....	71
Quadro 3.2 - Classificação dos itens conforme o conceito de usabilidade abordado.....	73
Quadro 3.3 - Composição da amostra segundo o tipo de produto comercializado pelo site....	74
Quadro 4.1 - Estatísticas descritivas dos itens: Correlação de Pearson e Bisserial.....	79
Quadro 4.2 - Estimação dos parâmetros dos itens.....	80
Quadro 4.3 - Estimação dos parâmetros com os itens selecionados.....	88
Quadro 4.4 - Descrição dos itens eliminados.....	89
Quadro 4.5 - Classificação dos itens âncora de acordo com os níveis do grau de usabilidade – na métrica (50,10).....	94
Quadro 4.6 - Descrição e classificação dos itens posicionados de nível 30 na escala gerada, de acordo com os aspectos de usabilidade associados.....	95
Quadro 4.7 - Descrição e classificação dos itens posicionados de nível 40 na escala gerada, de acordo com os aspectos de usabilidade associados.....	97
Quadro 4.8 - Descrição e classificação dos itens posicionados de nível 50 na escala gerada, de acordo com os aspectos de usabilidade associados.....	98
Quadro 4.9 - Descrição e classificação dos itens posicionados de nível 60 na escala gerada, de acordo com os aspectos de usabilidade associados.....	99
Quadro 4.10 - Distribuição dos itens de acordo com o conceito de usabilidade e os níveis na escala gerada.....	100
Quadro 4.11 – Itens do construto final e seus respectivos parâmetros e níveis na escala de usabilidade (50,10)	104

SUMÁRIO

1 INTRODUÇÃO	12
1.1 Contextualização.....	12
1.2 Problema de Pesquisa	13
1.3 Objetivos.....	15
1.3.1 Objetivo geral	15
1.3.2 Objetivos específicos	15
1.4 Justificativa	15
1.5 Limites do Trabalho.....	16
1.6 Estrutura do Trabalho	17
2 FUNDAMENTAÇÃO TEÓRICA.....	18
2.1 Usabilidade	18
2.1.1 Princípios ergonômicos para Interfaces Humano-Computador (IHC)	20
2.1.2 Técnicas de avaliação de usabilidade	31
2.1.2.1 Avaliação Heurística.....	32
2.1.2.2 Avaliação Analítica.....	33
2.1.2.3 Inspeção baseada em listas de verificação (Checklist)	33
2.1.2.4 Inspeção Cognitiva.....	34
2.1.2.5 Análise da Atividade.....	34
2.1.2.6 Grupo Focal	35
2.1.2.7 Avaliação Cooperativa.....	35
2.1.2.8 Teste de Usabilidade.....	36
2.1.2.9 Questionários de satisfação.....	37
2.2 Comércio Eletrônico	40
2.2.1 Conceito	40
2.2.2 Tipos de comércio eletrônico.....	40
2.2.3 O consumidor no comércio eletrônico.....	42
2.3 Escalas de Medidas	44
2.3.1 Tipologias de escalas	44
2.3.2 Vantagens e desvantagens das escalas.....	46
2.3.3 Elaboração de um conjunto de itens	47
2.4 Teoria da Resposta ao Item.....	49
2.4.1 Conceitos básicos.....	49
2.4.2 Questões críticas para a aplicação da Teoria de Resposta ao Item	54
2.4.3 A escolha do modelo mais adequado.....	56
2.4.4 Tamanho da amostra	57
2.4.5 Estimação	58
2.4.6 Softwares	59
2.5 Considerações Finais do Capítulo.....	60
3 PROCEDIMENTOS METODOLÓGICOS	62
3.1 Classificação da Pesquisa	62
3.2 Elaboração do Conjunto de Itens	63
3.2.1 Conceitos utilizados.....	64
3.2.1.1 Design da página.....	64
3.2.1.2 Navegação.....	66
3.2.1.3 Busca e informações	67
3.2.1.4 Suporte à tarefa de compra	69

3.2.2 Organização do conjunto de itens	69
3.2.3 Teste piloto	70
3.2.4 Itens resultantes	71
3.3 População e Amostra	73
3.4 Coleta de Dados	74
3.5 Tratamento dos Dados	76
4 ANÁLISE DOS RESULTADOS	78
4.1 Estatística Descritiva	78
4.2 Calibração dos Itens	80
4.3 Estimação das Habilidades	84
4.4 Construção da Escala de Medida do Grau de Usabilidade	86
4.4.1 Itens eliminados	88
4.4.2 Mudança de escala	92
4.4.3 Níveis âncora e itens âncora para a escala de usabilidade	93
4.4.4 Localização dos <i>sites</i> analisados na escala gerada	101
4.5 Considerações Finais do Capítulo	102
5 CONCLUSÕES E RECOMENDAÇÕES	105
5.1 Conclusões	105
5.2 Recomendações para Trabalhos Futuros	106
REFERÊNCIAS	108
APÊNDICE A – Endereço dos <i>sites</i> avaliados	118
APÊNDICE B - Conjunto de itens iniciais	121
APÊNDICE C – Curvas Características dos Itens	124
APÊNDICE D – Curva de Informação dos Itens	130
APÊNDICE E – Estimação dos parâmetros após a retirada de alguns itens.	136
APÊNDICE F – Curva de Informação dos conjuntos de itens	139

1 INTRODUÇÃO

1.1 Contextualização

A popularização da internet tem dinamizado significativamente a economia digital, e consequentemente impulsionado o comércio eletrônico. Em 2008 o comércio eletrônico no Brasil faturou 8,2 bilhões de reais, um crescimento nominal de 30% quando comparado ao ano de 2007 (E-BIT, 2009). Este crescimento indica que mais usuários têm feito uso deste serviço. Desta forma, dentro deste contexto surgem algumas questões referentes à gestão deste serviço, como por exemplo, a avaliação de desempenho das empresas de *e-commerce*, os fatores mais importantes na interação com o consumidor, como facilitar esta interação entre outros.

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; TROCCHIA; GWINNER, 2002; WANG; TANG; TANG, 2001; LONG; McMELLON, 2004; LIM; DUBINSKY, 2004; KIM; STOEL, 2004; SONG; ZINKHAN, 2003). 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.

Segundo Nielsen e Loranger (2006), a usabilidade em *websites* é um atributo de qualidade relacionado à facilidade de uso deste. Mais especificamente, (1) refere-se à rapidez com que os usuários podem aprender a usá-lo, (2) a eficiência deles ao usá-lo, (3) o quanto lembram deste, (4) o grau de propensão a erros e (5) o quanto gostam de utilizá-lo.

Winckler (2001) argumenta que problemas de usabilidade em site de *e-commerce* podem significar redução nas vendas ou mesmo venda nenhuma. Considerando-se que o grande aumento da quantidade de *sites* de *e-commerce* torna o mercado muito mais competitivo e globalizado, estratégias que priorizem o usuário e avaliem de forma coesa e objetiva a interação deste com o *website* são fundamentais para garantir competitividade por qualidade e inovação.

A Teoria da Resposta ao Item (TRI) é um conjunto de modelos matemáticos que pode ser empregada para se estabelecer medições, a partir de um conjunto de atributos relacionados com o que se pretende medir. Desta forma, é possível criar escalas de medida padronizadas (VARGAS, 2007). A TRI vem ganhando cada vez mais espaço no meio acadêmico por apresentar características importantes no tratamento de dados e na elaboração de escalas de medida. Hambleton, Swaminathan e Rogers (1991) citam que:

- As características dos itens são independentes da amostra de sujeitos utilizados;
- O escore do examinado independe do teste utilizado;
- O modelo é baseado nos itens individualmente e não no teste como um todo;
- Estabelece um modelo que não exige formas rigorosamente paralelas para avaliar a fidedignidade.

1.2 Problema de Pesquisa

Para avaliar o desempenho de qualquer sistema, é necessário compreender perfeitamente as estruturas e relacionamento dos elementos envolvidos neste (BITITCI, 1995; LEBRAS, 1995; LOHAMAN; FORTUIN; WOUTERS, 2004). Entretanto, a compreensão destas estruturas e relacionamentos depende da precisão e exatidão na medição destes elementos, seja ela em sistemas reais ou virtuais.

Atualmente, a economia digital vem ganhando mais espaço na economia mundial, e dentro desta, destacam-se os sistemas virtuais, como o e-commerce B2C. Este crescimento indica que cada vez mais negócios vêm sendo realizados via internet, mas pouco se têm falado de medidas objetivas dos elementos envolvidos neste cenário (USHIWATA ET AL, 2006).

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, 1997, 2002, 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).

Nielsen e Loranger (2006) em seus estudos criticam os métodos subjetivos de avaliação de satisfação dos usuários argumentando que estes são pouco informativos e confiáveis porque os usuários tendem a atribuir avaliações generosas mesmo quando têm grandes dificuldades ao utilizar um site. Uma razão disso é o desejo humano geral de ser gentil e aceito. Outra razão, segundo os autores é que os usuários frequentemente não têm conhecimento do seu desempenho ao testar um site. Se encontrarem informações para seus problemas, eles acreditam que o site foi útil, e não percebem que esse site poderia conter informações muito mais relevantes que não foram prontamente disponibilizadas para eles.

A subjetividade e a falta de sistematização nos resultados dificultam a comparabilidade entre os sistemas e a identificação das características mais importantes. A dificuldade na comparabilidade dos desempenhos de usabilidade impede o acompanhamento da evolução do sistema e seu posicionamento perante seus concorrentes. Já a dificuldade em identificar claramente as características pode fazer com que recursos sejam alocados em características pouco importantes enquanto características de fato importantes sejam negligenciadas.

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 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 de usabilidade.

Neste sentido, tendo como intuito a geração de entendimento sobre as estruturas e relacionamento dos elementos envolvidos nas questões referentes à usabilidade em *e-commerce*, resume-se a problemática do estudo com a seguinte questão:

Como realizar a avaliação de usabilidade em *sites* de *e-commerce* utilizando a Teoria da Resposta ao Item?

1.3 Objetivos

1.3.1 Objetivo geral

Elaborar um construto para medir usabilidade em *sites* de *e-commerce* utilizando a Teoria da Resposta ao Item

1.3.2 Objetivos específicos

- Identificar aspectos envolvendo a produtividade na Interação Humano-Computador (IHC) em *sites* de *e-commerce*;
- Desenvolver um conjunto de itens para avaliar usabilidade em *sites* de *e-commerce*;
- Calibrar os itens no contexto de *sites* de *e-commerce*;
- Criar uma escala padronizada para medir usabilidade em *sites* de *e-commerce* com suporte da Teoria da Resposta ao Item.

1.4 Justificativa

Segundo e-bit (2008), no primeiro semestre de 2008 o índice de usuários insatisfeitos com as lojas virtuais foi de 13,52% o que representa um pequeno aumento em relação a 2007, onde a insatisfação era de 13%. Esta redução da satisfação está relacionada ao fato de que o consumidor tem se tornado mais exigente com todo o processo de negociação virtual desde a interação inicial com o site até o pós venda. A pesquisa afirma ainda que os requisitos que mais fazem o índice de satisfação do consumidor cair são: descumprimento no prazo de entrega, baixa qualidade no atendimento e deficiência na usabilidade de navegação.

A importância da usabilidade em *sites* comerciais é reforçada por Nielsen e Loranger (2006), onde afirmam que cerca de 34% das vezes em que um usuário realiza uma tarefa em um site de *e-commerce* ele não consegue completar sua tarefa com sucesso. O que, por se tratar de objetivos na maioria das vezes comercial, pode significar cliente perdido.

Sendo assim, apresentar uma boa usabilidade em um site comercial pode representar um importante diferencial competitivo. Grande parte das avaliações de usabilidade existentes na literatura é feita de forma subjetiva, baseada em heurísticas desenvolvidas por especialistas ou

em respostas subjetivas a questionários sobre satisfação do usuário, o que causam certa incerteza na sistematização das avaliações (CYBIS, 2007; NIELSEN; LORANGER, 2006)

Desta forma, é importante realizar uma boa avaliação de usabilidade, objetiva e sistemática, capazes de gerar entendimento, comparabilidade e confiabilidade à estrutura de avaliação e aos resultados. Para isso, é proposta a elaboração de um construto baseado em conceitos de usabilidade em e-commerce. E sua validação é feita através de uma ferramenta matemática robusta – Teoria da Resposta ao Item - capaz de gerar sistematização nos resultados e comparabilidade por meio da criação de uma escala de medida.

Este tipo de abordagem não é encontrado na literatura de avaliação de usabilidade, representando assim uma relevância teórica importante para o tema.

1.5 Limites do Trabalho

O embasamento teórico do presente estudo limita-se ao campo de usabilidade em *sites* de *e-commerce* e prioriza uma avaliação objetiva e quantitativa.

Dentro do contexto de *e-commerce* o estudo limitou-se ao mercado *business to consumer* (B2C).

O estudo não focou em aspectos relacionados a teste com usuários, uma vez que o objetivo do construto não é avaliar o usuário e sim o *website*.

A avaliação dos *websites* foi baseada em apenas um usuário, o que limitou a abordagem a aspectos objetivos.

A quantidade de itens e de *sites* analisados esteve condicionada ao tempo disponível para realização da pesquisa.

O construto proposto não se destina a substituir uma avaliação completa de usabilidade, que envolve a realização de ensaios de interação com usuários e a participação de especialistas em interfaces.

1.6 Estrutura do Trabalho

Este trabalho está estruturado em cinco capítulos. No primeiro é realizada uma introdução ao assunto, contextualizando o tema, discutindo-se a problemática da pesquisa, expondo os objetivos, a justificativa e as limitações do trabalho.

No capítulo dois é apresentada a fundamentação teórica sobre os temas que sustentam a pesquisa: Usabilidade, Comércio Eletrônico, Escalas de Medidas e Teoria da Resposta ao Item.

No capítulo três são detalhados os procedimentos metodológicos referentes à elaboração do questionário, os aspectos considerados, a organização do conjunto de itens, a amostragem, coleta e tratamento dos dados, o teste piloto e os itens resultantes.

No capítulo quatro é apresentada a avaliação e discussão dos resultados juntamente com a criação da escala proposta.

No capítulo cinco são apresentadas as conclusões e as recomendações para trabalhos futuros, seguidas das referências bibliográficas utilizadas e dos apêndices.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo aborda os conceitos básicos de usabilidade, comércio eletrônico, escalas de medidas e Teoria da Resposta ao Item (TRI). Dentro da abordagem sobre usabilidade são levantadas considerações sobre as definições encontradas na literatura juntamente com os critérios ergonômicos envolvidos. Algumas das técnicas de avaliação de usabilidade também são discutidas. Além dos conceitos de comércio eletrônico são levantados também os seus tipos e sua relação com o usuário e a usabilidade. Na seção destinada aos conceitos de escalas de medida são abordadas também suas tipologias, vantagens e desvantagens inseridas no contexto de usabilidade e *e-commerce*. E por fim, a conceituação da Teoria da Resposta ao Item é feita juntamente com um levantamento histórico desta ferramenta, alguns de seus modelos matemáticos e questões críticas para sua aplicação.

2.1 Usabilidade

Com o avanço da tecnologia e consequentemente a popularização dos computadores, a partir da década de 1980, este passou a ser utilizado por pessoas das mais diversas áreas do conhecimento, desde técnicos especializados em computação até leigos em informática básica. Desta forma, surgiu a necessidade de criar interfaces gráficas simples e fáceis de usar surgindo então a engenharia de usabilidade.

Segundo Mayhew (1999), a engenharia de usabilidade é uma disciplina que fornece métodos estruturados para obtenção de usabilidade em projetos de interface com o usuário durante o desenvolvimento do produto ou *website*. Esta é uma disciplina com raízes em várias outras disciplinas básicas, incluindo (i) psicologia cognitiva, que estuda as percepções humanas (visão, audição etc) e a cognição (memória, aprendizagem, resolução de problemas, tomada de decisão, reações, etc); (ii) psicologia experimental, que utiliza métodos empíricos para a medida de desempenho do usuário e de sua satisfação; (iii) etnografia, que estuda aspectos culturais para a descrição das características dos usuários; (iv) ergonomia, que estuda os esforços físicos, cognitivos e organizacionais para maximizar o desempenho e reduzir a fadiga e o desconforto na utilização de sistemas; (v) engenharia de software, que relaciona a

usabilidade com o ciclo de vida dos sistemas; entre outros. Configurando assim uma estrutura centrada no usuário.

A norma ISO 9241 define usabilidade como a capacidade que um sistema interativo oferece a seu usuário, em determinado contexto de operação, para realização de tarefas de maneira eficaz, eficiente e agradável (ISO 9241-11, 1998). Seguindo esta definição, Cybis (2007) classifica a usabilidade como uma qualidade que caracteriza o uso dos programas e aplicações. Isto significa que depende do contexto e dos objetivos do usuário, não sendo portanto, uma qualidade intrínseca de um sistema. A Figura 2.1 demonstra mais claramente a estrutura da usabilidade, onde se pode verificar a necessidade de identificar os objetivos do sistema, a descrição dos componentes do contexto de uso (usuário, tarefa, equipamento e ambiente) e a decomposição da usabilidade em atributos passíveis de serem verificados e mensurados (eficácia, eficiência e satisfação). O contexto de uso pode ser a descrição de um contexto já existente ou a especificação de um pretendido, ou seja, dependente dos objetivos. Este contexto define o resultado de uso, portanto para um perfeito entendimento da interação do usuário com o sistema, este deve ser detalhado o máximo possível para identificar possíveis problemas de usabilidade.

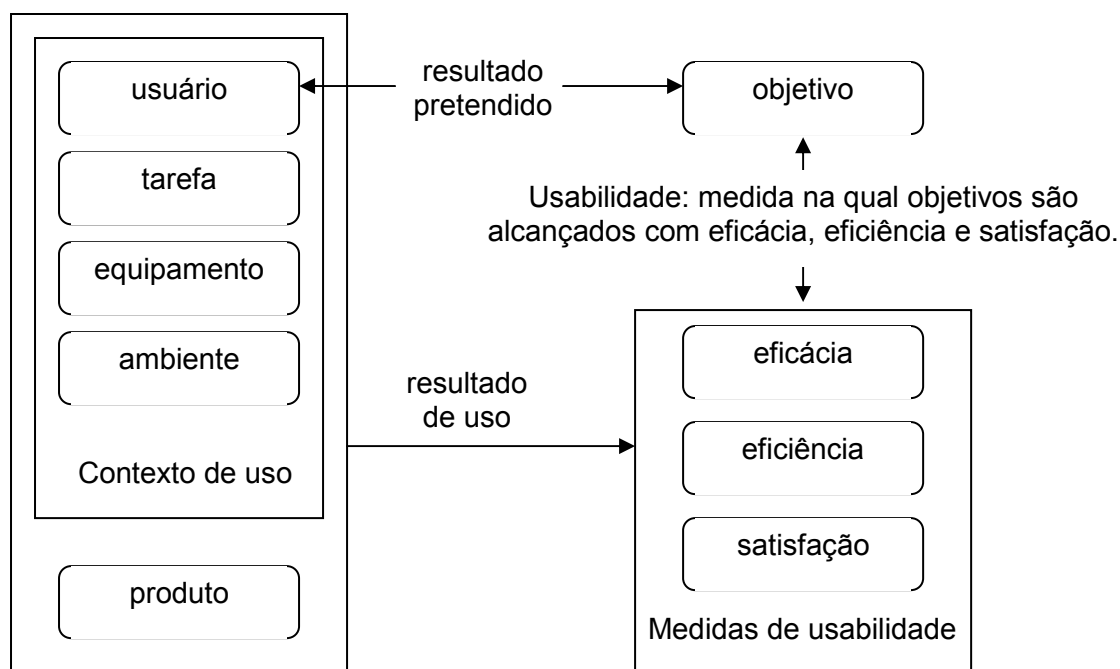


Figura 2.1 - Estrutura da usabilidade.
Fonte: Adaptado de ISO 9241-11 (1998).

As medidas de usabilidade dependem diretamente dos objetivos e do resultado do uso. O presente trabalho concentra-se na esfera de medidas de usabilidade, mais especificamente na

parte de eficácia e eficiência, considerando os objetivos de sistemas de comércio eletrônico e o contexto de uso referente a tarefa e ao ambiente.

A eficácia é o grau de precisão e de abrangência obtidos na interação do usuário com o sistema, e relaciona-se com a quantidade de recursos despendidos para alcançar os objetivos desejados. A medição deste atributo é dependente do fato do usuário ou consumidor conseguir realmente concluir a tarefa desejada, ou seja, se ele encontrar as informações desejadas ou comprar um produto. A eficiência está relacionada ao emprego de recursos aplicados para o usuário atingir seus objetivos, como tempo, esforços mentais, físicos, operacionais, ambientais, de hardware, software, etc. A satisfação pode ser compreendida como o grau de conforto que o usuário tem ao utilizar o software. A maneira mais comum para medir a satisfação do usuário é via questionário de satisfação (HIX; HARTSON, 1993; JORDAN 1998; ISO 9241-11, 1998). A maioria das definições ou descrições de usabilidade inclui uma vertente de aprendizagem (EGGER 1999; MAYHEW 1999). Embora existam diferentes terminologias utilizadas por vários autores, na sua essência, as definições de usabilidade são as mesmas. No presente estudo adotou-se a definição da ISO 9241-11 (1998).

2.1.1 Princípios ergonômicos para Interfaces Humano-Computador (IHC)

O presente trabalho tem como propósito trabalhar a esfera de medidas de usabilidade, demonstrada na Figura 2.1. Entretanto, é necessário conhecer a estrutura relacionada ao contexto de uso, uma vez que a dinâmica neste contexto representa o resultado de uso e pode desta forma, indicar a origem de eventuais problemas de usabilidade. Além de possibilitar o entendimento nas relações de eficácia, eficiência e satisfação como um sistema de causa-e-efeito. O aprofundamento no contexto de uso está relacionado com princípios ergonômicos de Interfaces Humano-Computador (IHC), que procuram nortear a qualidade da interação entre o usuário e o sistema.

Entretanto, não existe consenso sobre o conjunto ideal de atributos de usabilidade que devem ser considerados no desenvolvimento de uma interface, seja ela específica para *sites de e-commerce* ou qualquer outro propósito. Diversos autores e normas apresentam conjuntos de atributos considerados essenciais para o desenvolvimento de uma interface homem-computador de qualidade (NIELSEN, 1993; BASTIEN; SCAPIN, 1993; DIX ET AL, 1998; HFDG, 2001; SHNEIDERMAN; PLAISANT, 2004; ISO 9241-11, 1998; ISO 9241-110, 2006; HHS, 2006). Existe porém, uma “configuração de base” a partir da qual uma interface

pode favorecer o estabelecimento da usabilidade na relação usuário-sistema. Esta configuração é construída com base nestes diversos atributos chamados de critérios ergonômicos, princípios ergonômicos ou heurísticas de usabilidade (CYBIS, 2007).

Dentro destas “configurações de base” as mais comuns na literatura estão relacionadas à condução, controle e adaptação à tarefa, consistência, gestão de erros e carga de trabalho (NIELSEN, 1993; BASTIEN; SCAPIN, 1993; DIX ET AL, 1998; HFDG, 2001; SHNEIDERMAN; PLAISANT, 2004; ISO 9241-110, 2006, CYBIS, 2007).

2.1.1.1 Condução

Segundo Bastien e Scapin (1993), a condução em um sistema interativo está relacionada aos mecanismos disponíveis para advertir, orientar, informar, instruir e guiar os usuários durante suas interações com o computador. Para estes autores, a condução é subdividida em presteza (permitem ao usuário identificar o estado ou contexto no qual se encontra, assim como as ferramentas de ajuda à sua acessibilidade), agrupamento e distinção de itens (diz respeito à organização visual dos itens de informação relacionados uns com os outros de alguma maneira), *feedback* imediato (respostas do sistema às ações do usuário) e legibilidade.

Para Dix *et al* (1998) a percepção do estado interno de um sistema a partir da sua representação observável, e o *feedback* imediato agregam ao sistema robustez. A questão da realimentação continua durante uma interação (*feedback*) é discutida também por Nielsen (1993), Shneiderman (1987), HFDG (2001) e ISO 9241:110 (2006). Para Nielsen (1993), o sistema deve sempre manter o usuário informado sobre o que está acontecendo através de um *feedback* rápido e apropriado. O usuário deve ser mantido informado sobre mudanças no status durante a interação com o sistema.

Uma boa condução também deve conduzir o usuário a sair rapidamente de alguma parte que não lhe interessa, sem comprometer seu caminho interativo, ou seja, o sistema não deve conter armadilhas ao usuário (NIELSEN, 1993). Quando o usuário estiver interagindo com uma sequência de diálogos, o sistema deve ser capaz de indicar claramente o final do diálogo (SHNEIDERMAN, 1987).

Para Bastien e Scapin (1993) agrupamento e distinção de itens diz respeito à organização dos itens de uma classe. A compreensão de uma tela pelo usuário depende, dentre outras

coisas, da ordenação, do posicionamento, e da distinção dos objetos (imagens, textos, comandos, etc.) que são apresentados.

A legibilidade no contexto de Interação Humano-Computador é conceituada como as características lexicais das informações apresentadas na tela que possam dificultar ou facilitar a leitura desta informação (brilho, contraste letra/fundo, tamanho da fonte, espaçamento entre palavras, espaçamento entre linhas, espaçamento de parágrafos, comprimento da linha, etc.) (BASTIEN; SCAPIN, 1993). A legibilidade é discutida também por HFDG (2001) e Cybis (2007). Para este último, a legibilidade é um critério importante quando a tarefa é de leitura e o público-alvo inclui pessoas idosas ou com problemas de visão. HFDG (2001) descreve algumas recomendações para estruturação de textos com maior legibilidade, entre elas:

- O espaçamento entre os caracteres deve ser pelo menos 10% do maior caractere;
- O espaçamento entre palavras deve ser de pelo menos um caractere;
- Todos os parágrafos devem ser separados por um espaço em branco;
- O contraste do caractere com o fundo deve estar entre 6:1 e 10:1;
- O número de fontes em cada display não deve ser maior que dois;
- Os textos devem combinar letras maiúsculas e minúsculas;
- Palavras em letra maiúscula devem ser utilizadas somente em títulos, frases chaves e itens curtos para chamar a atenção do usuário para algo realmente importante.
- Palavras sublinhadas devem ser evitadas.

Bastien e Scapin (1993) também descrevem algumas recomendações para melhorar a condução em uma interação, entre elas pode-se citar:

- Para cada campo de dados fornecer um rótulo.
- Indicar o tamanho do campo quando ele é limitado.
- Quando necessário, fornecer no rótulo informações suplementares.
- Dar um título a cada janela.
- Organizar os itens em listas hierárquicas
- Organizar as opções de um diálogo por menus, em função dos objetos aos quais elas se aplicam.
- Quando várias opções são apresentadas, sua organização deve ser lógica, a organização deve representar uma organização funcional relevante ou significativa (ordem alfabética, frequência de uso, etc).

- Fazer uma distinção visual clara de áreas que têm diferentes funções (área de comandos, área de mensagens, etc).
- Rótulos devem estar em letras maiúsculas.
- Exibir texto contínuo em colunas largas, de ao menos 50 caracteres por linha.

2.1.1.2 Controle e adaptação da tarefa

Um dialogo é controlável quando o usuário está apto para iniciar e controlar a direção e o ritmo da interação, até o ponto em que o objetivo for alcançado. E um sistema interativo é adaptável à tarefa quando este dá suporte ao usuário para completar sua tarefa, por exemplo, quando a funcionalidade e o dialogo são baseados nas características da tarefa. O dialogo deve apresentar ao usuário informações relacionadas para o sucesso na efetivação da tarefa, ou seja, as necessidades da tarefa determinam a qualidade, a quantidade e o tipo de informação a ser apresentada (ISO 9241-110, 2006).

Quando os usuários definem explicitamente suas entradas, e quando estas entradas estão sob o controle deles, erros e ambiguidades são limitados. Além disso, o sistema será mais bem aceito pelos usuários se eles tiverem controle sobre o diálogo (BASTIEN e SCAPIN, 1993). Para Dix et al. (1998) deve haver transferência de controle entre usuário e sistema, em função do contexto, principalmente em aplicações de automação crítica, onde esta característica ajuda no balanceamento de atividades de operação.

Bastien e Scapin (1993) subdividem o controle explícito em outras duas características: ações explícitas do usuário e controle do usuário. A primeira refere-se às relações entre o processamento pelo computador e as ações do usuário. O sistema deve processar somente aquelas ações solicitadas pelo usuário e somente quando solicitado a fazê-lo. Já o controle do usuário, refere-se ao fato de que os usuários deveriam estar sempre no controle do processamento do sistema, por exemplo: interromper, cancelar, suspender e continuar. Cada ação possível do usuário deve ser antecipada e opções apropriadas devem ser oferecidas.

Nielsen (1993) e Shneiderman (1987) tratam da questão relativa à adaptação do sistema ao usuário quando recomendam o fornecimento de atalhos que além de contribuir para o controle do usuário pode agilizar a interação, principalmente de usuários mais experientes, que não têm a mesma necessidade informacional que os novatos.

Para Bastien e Scapin (1993), uma interface não pode atender ao mesmo tempo a todos os seus usuários em potencial. Para que ela não tenha efeitos negativos sobre o usuário, esta interface deve, conforme o contexto, se adaptar ao usuário. Por outro lado, quanto mais variadas forem as maneiras de realizar uma tarefa, maiores são as chances que o usuário possui de escolher e dominar uma delas no curso de seu aprendizado. Deve-se portando fornecer ao usuário procedimentos, opções, comandos diferentes permitindo-lhe alcançar um mesmo objetivo. Os autores descrevem também algumas recomendações referentes ao controle e adaptação da tarefa:

- A sequência de entrada de dados deve poder ser modificada para se adaptar a ordem preferida pelo usuário.
- Quando o formato de um texto não puder ser previsto com antecedência, deve-se proporcionar ao usuário os meios para definir e salvar os formatos que ele venha a precisar.
- O usuário deve poder definir os nomes dos campos de dados que ele(a) venha a criar.
- Sempre faça necessário que o usuário tecle um ENTER explícito para iniciar o processamento de dados digitados; não inicie um processamento (por exemplo, atualizar um arquivo) como efeito colateral de outra ação (por exemplo, imprimir um arquivo).
- Se a seleção do menu é feita através de dispositivo de apontamento, faça a ativação em dois passos, onde a primeira ação designa a opção selecionada e uma segunda ação distinta faz uma entrada de controle explícita.
- Entradas de comandos do usuário devem ser seguidas de um ENTER depois de editadas.
- Deixar ao usuário o controle do ritmo de suas entradas de dados, e não pelo computador ou por eventos externos.
- O cursor não deve ser automaticamente movido sem o controle do usuário (com exceção de procedimentos estáveis e bem conhecidos como o preenchimento de formulários).
- Possibilitar aos usuários interromper ou cancelar a transação ou processo atual.
- Fornecer uma opção CANCELAR a qual tem o efeito de apagar qualquer mudança que acabou de ser feita e trazer a tela para seu estado anterior.
- Prever atalhos. Permitir que usuários experientes contornem uma série de seleções por menu através da especificação de comandos ou de atalhos de teclado.

- Prever a escolha de entradas simples ou múltiplas conforme a experiência do usuário.
- Autorizar diferentes modos de diálogo correspondentes aos diferentes grupos de usuários (ex. prever uma presteza adaptada ao nível de experiência do usuário).
- Fornecer um tutorial passo a passo para os usuários novatos.

2.1.1.3 Consistência

Consistência é um dos mais básicos princípios de usabilidade. Se os usuários sabem que o mesmo comando ou a mesma ação irá sempre ter o mesmo efeito, eles se sentirão mais confiante em utilizar o sistema, e eles serão encorajados a tentar explorar o sistema. A mesma informação deve ser representada na mesma localização em todas as telas e as caixas de diálogos devem ser formatadas da mesma maneira para facilitar o reconhecimento por parte do usuário (NIELSEN, 1993).

HFDG (2001) argumenta que a consistência deve ser priorizada no design da página, no alinhamento de textos, no formato dos rótulos, campos, mensagens e instruções, em informações gráficas, nas cores empregadas no sistema juntamente com seus códigos e nas respostas às questões durante a interação.

Para Dix *et al* (1998), a consistência facilita a aprendizagem, por trabalhar com a semelhança de comportamento em tarefas com objetivos similares, além de desenvolver o atributo da previsibilidade do sistema (determinação do comportamento futuro do sistema com base no histórico; visibilidade das operações realizadas).

Bastien e Scapin (1993) referem-se à consistência (ou homogeneidade) como a forma na qual as escolhas na concepção da interface (códigos, denominações, formatos, procedimentos, etc.) são conservadas idênticas em contextos idênticos, e diferentes para contextos diferentes. Os procedimentos, rótulos, comandos, entre outros, são melhor reconhecidos, localizados e utilizados, quando seu formato, localização, ou sintaxe são estáveis de uma tela para outra, de uma seção para outra. Nestas condições o sistema é mais previsível e a aprendizagem mais generalizável; os erros são diminuídos. Segundo estes, a falta de homogeneidade nos menus, por exemplo, pode aumentar consideravelmente os tempos de procura. A falta de homogeneidade é também uma razão importante da recusa na utilização. Estes autores também consideram como um critério ergonômico o significado dos códigos e denominações, que diz respeito à adequação entre o objeto ou a informação apresentada ou pedida, e sua referência. Quando a codificação é significativa, a recordação e o reconhecimento são

melhores. Códigos e denominações não significativos para os usuários podem lhes sugerir operações inadequadas para o contexto, lhes conduzindo a cometer erros.

Exemplos de recomendações, segundo Bastien e Scapin (1993):

- Localização similar dos títulos das janelas.
- Formatos de telas semelhantes.
- Procedimentos similares de acesso às opções dos menus.
- Na condução, sempre utilizar as mesmas pontuações e as mesmas construções de frases.
- Apresentar na mesma posição os convites (prompts) para as entrada de dados ou de comandos.
- Os formatos dos campos de entrada de dados devem sempre ser os mesmos.
- O título deve transmitir o que ele representa e ser distinto de outros títulos;
- Explicitar as regras de contração ou de abreviação;
- Utilizar códigos e denominações significativas e familiares em vez de códigos e denominações arbitrárias (ex.: utilizar M para masculino e F para feminino em vez de 1 e 2).

2.1.1.4 Gestão de erros

A gestão de erros diz respeito a todos os mecanismos que permitem evitar ou reduzir a ocorrência de erros, e quando eles ocorrem que favoreçam sua correção. Os erros são aqui considerados como entrada de dados incorretos, entradas com formatos inadequados, entradas de comandos com sintaxes incorretas, etc. A gestão de erros pode ser subdividida em três sub-critérios: proteção contra os erros, correção de erros e qualidade das mensagens de erro. O primeiro diz respeito aos mecanismos empregados para detectar e prevenir os erros de entradas de dados, comandos, possíveis ações de consequências desastrosas e/ou não recuperáveis. É preferível detectar os erros no momento da digitação do que no momento da validação. A correção dos erros, diz respeito aos meios colocados a disposição do usuário com o objetivo de permitir a correção de seus erros. Os erros são bem menos perturbadores quando eles são fáceis de corrigir. A qualidade das mensagens de erro refere-se à pertinência, a legibilidade e a exatidão da informação dada ao usuário sobre a natureza do erro cometido

(sintaxe, formato, etc.) e sobre as ações a executar para corrigi-lo (BASTIEN; SCAPIN, 1993).

Para Nielsen (1993), melhor do que ter boas mensagens de erro é evitar o erro. Há muitas situações que são conhecidas por serem suscetíveis a erros (NORMAN 1983; REASON, 1990; SENDERS; MORAY 1991), e sistemas podem ser concebido para evitar a colocação do usuário em tais situações. Por exemplo, toda vez que o usuário é solicitado a enunciar algo, existe um risco de erros ortográficos, então, selecionar um arquivo a partir de um menu em vez de escrever é uma opção de redesenho do sistema para eliminar toda uma categoria de erros. Erros com consequências especialmente graves, também podem ser reduzidos em frequência, solicitando-se ao usuário a reconfirmação da ação antes de ir em frente (NIELSEN, 1993).

O sistema interativo deve auxiliar o usuário na detecção de erros, além de evitá-los. Por exemplo, em aplicações de e-commerce, no preenchimento de formulários, a utilização da pontuação deve, sempre que possível, ser automática, evitando erros de entrada de dados. E quando um erro ocorrer, uma explicação de fácil entendimento pelo usuário, deve ser fornecida (ISO 9241-110, 2006).

A prevenção e manipulação simplificada de erros também são abordadas em Shneiderman (1987) e HFDG (2001). Para este último, o usuário deve estar apto a parar uma ação a qualquer momento para corrigir eventuais erros. O sistema deve ser capaz de distinguir erros de programação, de hardware ou de operação, e o alerta de erro deve permanecer visível ao usuário até que ele responda com a ação mais apropriada. Algumas das recomendações de HFDG (2001) e Bastien e Scapin (1993) podem ser descritas como:

- Quando o usuário termina uma seção e existe o risco de perda os dados, deve haver uma mensagem lhe avisando deste fato e lhe pedindo confirmação do final da seção.
- Os rótulos dos campos devem estar protegidos (não devem ser acessíveis ao usuário).
- As apresentações que acompanham as entrada de dados devem estar protegidas. Os usuários não podem modificar as informações contidas nestes campos.
- Depois de um erro de digitação de um comando ou de dados, dar ao usuário a possibilidade de corrigir somente a parte dos dados ou do comando que está errado.
- Todas as ações possíveis sobre uma interface devem ser consideradas e mais particularmente as digitações acidentais a fim de que entradas não esperadas sejam detectadas.

- Quando o usuário entrar com um formato de data incorreto, a mensagem de erro deve indicar: “formato correto MM/DD/YY.
- Utilizar os termos tão específicos quanto possíveis para as mensagens de erros.
- Utilizar mensagens de erro tão breves quanto possível.
- Adotar um vocabulário neutro, não personalizado, não repreensivo nas mensagens de erro; evitar o humor.
- Fornecer a possibilidade de modificar os comandos no momento de sua digitação.
- Se o usuário não percebe que cometeu um erro de digitação, lhe dar a possibilidade de efetuar, no momento da detecção do erro, as correções apropriadas.

2.1.1.5 Carga de trabalho

Este critério está ligado a um contexto de trabalho intenso e repetitivo e diz respeito a todos os elementos da interface que têm um papel importante na redução da carga cognitiva e perceptiva do usuário, e no aumento da eficiência do diálogo. Quanto maior for a carga de trabalho, maior será a probabilidade de cometer erros. E também, quanto menos o usuário for distraído por informação desnecessária, mais ele será capaz de desempenhar suas tarefas eficientemente. Além disso, quanto menos ações são necessárias, mais rápidas serão as interações. A tarefa cognitiva imposta ao usuário deve ser breve, concisa, com o mínimo possível de ações e com baixa densidade informacional (BASTIEN; SCAPIN, 1993).

O usuário não deve ter de relembrar informações de uma parte de um diálogo em outra. Instruções de uso do sistema devem ser visíveis ou facilmente revisáveis sempre que apropriadas (NIELSEN, 1993).

Para HFDG (2001), uma janela deve conter todas as informações relevantes e deve permitir que o usuário complete a tarefa sem ter que lembrar de alguma informação adicional, minimizando assim a carga cognitiva do usuário. Por exemplo, o sistema pode fornecer um calendário se a tarefa exigir a entrada de algum dado referente a datas. A carga cognitiva também pode ser reduzida com a utilização correta de cores e seus significados.

Bastien e Scapin (1993) e HFDG (2001) recomendam algumas ações para reduzir a carga mental em uma interação:

- Para dados numéricos, a entrada de zeros à esquerda não deve ser necessária.

- O número de mnemônicos, códigos especiais ou longas sequências, e instruções especiais que o usuário necessitar se esforçar para aprender devem ser minimizados.
- Se códigos forem mais longos que 4 ou 5 caracteres, use mnemônicos ou abreviaturas.
- Permitir ao usuário entradas de dados sucintas.
- Quando uma unidade de medida está associada a um campo, inclua a unidade como parte do campo de dados, ao invés de fazer o usuário digitá-la.
- Minimize o número de passos necessários para se fazer uma seleção em menu.
- Não faça o usuário entrar com dados que poderiam ser derivados pelo computador.
- Para entrada de dados, exiba os valores *default* atuais nos campos apropriados.
- Fornecer somente dados que sejam necessários e imediatamente usáveis para qualquer transação.
- Os dados não devem requerer transformação de unidades.
- A linguagem de consulta deve usar o mínimo de quantificadores em sua formulação.
- Não fazer com que os usuários precisem lembrar de dados exatos de uma tela para outra.

Baseado nas descrições destas “características base” expostas, é possível estabelecer algumas relações entre aspectos do contexto de operação dos sistemas, e os critérios ergonômicos a priorizar em uma atividade de projeto e/ou avaliação de usabilidade (CYBIS, 2007).

O Quadro 2.1 demonstra os critérios considerados por cada um dos autores pesquisados, de acordo com as características mais comuns entre eles, as chamadas “características base”.

Autor	Característica base				
	Condução	Controle e adaptação da tarefa	Consistência	Gestão de erros	Carga de trabalho
Shneiderman (1987)	<ul style="list-style-type: none"> - Fornecer <i>feedback</i> informativo; - Marcar o final dos diálogos. 	<ul style="list-style-type: none"> - Fornecer atalhos; - Permitir o cancelamento das ações; - Fornecer controle e iniciativa ao usuário. 	<ul style="list-style-type: none"> - Perseguir a consistência. 	<ul style="list-style-type: none"> - Prevenção e manipulação simples de erros. 	<ul style="list-style-type: none"> - reduzir a carga de memória de trabalho.
Bastien e Scapin (1993)	<ul style="list-style-type: none"> - Presteza; - Agrupamento e distinção entre itens; - Feedback; - Legibilidade. 	<ul style="list-style-type: none"> - Ações explícitas; - Controle do usuário; - Flexibilidade; - Consideração da experiência do usuário. 	<ul style="list-style-type: none"> - Homogeneidade; - Significado de códigos e denominações. 	<ul style="list-style-type: none"> - Proteção contra erros; - Qualidade das mensagens de erro; - Correção de erros. 	<ul style="list-style-type: none"> - Brevidade; - Densidade informacional; - Compatibilidade.
Nielsen (1993)	<ul style="list-style-type: none"> - Feedback; - Indique claramente a saída; - Ajuda e documentação. 	<ul style="list-style-type: none"> - Prover atalhos; - Falar a linguagem do usuário. 	<ul style="list-style-type: none"> - Manter a consistência. 	<ul style="list-style-type: none"> - Prevenção de erros; - Prover boa mensagem de erro. 	<ul style="list-style-type: none"> - Diálogo simples e natural; - Minimizar a carga de memória do usuário.
Dix et al (1998)	<ul style="list-style-type: none"> - Familiaridade; - Observabilidade; - Capacidade de resposta. 	<ul style="list-style-type: none"> - Iniciativa de diálogo; - Paralelismo; - Capacidade de migração; - Capacidade de configuração; - Conformidade a tarefa. 	<ul style="list-style-type: none"> - Previsibilidade; - Capacidade de síntese; - Capacidade de generalização. 	<ul style="list-style-type: none"> - Capacidade de recuperação. 	<ul style="list-style-type: none"> - Capacidade de substituição.
HFDG (2001)	<ul style="list-style-type: none"> - Estrutura dos textos; - Feedback; - Apresentação simplificada e bem organizada das informações; - Não esconder informações; - Fornecer ajuda on-line. 	<ul style="list-style-type: none"> - Facilitar a navegação; - Fornecer controle sobre mapas e display táctico. 	<ul style="list-style-type: none"> - Prover consistência na paginação, listas, tabelas e informações gráficas; - Cores e seus significados e posições. 	<ul style="list-style-type: none"> - Gerenciamento de erros. 	<ul style="list-style-type: none"> - Minimizar a carga de memória do usuário; - Minimizar a densidade informacional.
ISO 9241-110 (2006)	<ul style="list-style-type: none"> - Auto descrição 	<ul style="list-style-type: none"> - Adaptação à tarefa; - Adequação para aprendizagem; - Controle ao usuário. - Facilidade de individualização. 	<ul style="list-style-type: none"> - Conformidade às expectativas do usuário; 	<ul style="list-style-type: none"> - Tolerância a erros 	<ul style="list-style-type: none"> -

Quadro 2.1 – Critérios ergonômicos considerados pelos autores segundo as características base.

Esta subseção levantou algumas das definições e recomendações ergonômicas para projeto de interface humano-computador com maior usabilidade. Entretanto, segundo Cybis (2007), tanto nos critérios ergonômicos quanto nas recomendações expostas, podem haver afirmações conflitantes, sendo necessário, examinar os elementos do contexto de uso esperado para o sistema, de modo a identificar os requisitos do projeto (intuitivamente em função de usuários novatos, acessibilidade em função de usuários idosos, etc) e a recomendação ergonômica que seja capaz de responder a tal requisito.

Na próxima seção estes critérios serão sistematizados em metodologias variadas para auxiliarem na avaliação da usabilidade.

2.1.2 Técnicas de 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) e que é ilustrado pela Figura 2.1. Sendo assim, verifica-se a existência de três características fins, que podem determinar a avaliação de um sistema ou de uma interação de forma objetiva ou subjetiva. São características que dependem diretamente da tarefa a ser realizada, do ambiente, do equipamento e do 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.

No contexto geral, a avaliação de usabilidade consiste na verificação do desempenho de uma interface e identificação de eventuais problemas de ergonomia de interface ou de usabilidade.

Um problema de ergonomia é identificado quando um aspecto de interface está em desacordo com as características dos usuários e da maneira pela qual ele realiza sua tarefa. Já um problema de usabilidade é observado em determinadas circunstâncias, quando uma característica do sistema interativo (problema de ergonomia) ocasiona a perda de tempo, compromete a qualidade da tarefa ou mesmo inviabiliza sua realização. Como consequência, este problema estará aborrecendo, constrangendo ou até traumatizando a pessoa que utiliza o sistema interativo (CYBIS, 2007). Desta forma, observa-se que os efeitos de um problema de usabilidade se fazem sentir diretamente sobre o usuário e indiretamente sobre sua tarefa.

Assim, por exemplo, a sobrecarga perceptiva (devido a dificuldade de leitura), cognitiva (devido à desorientação) ou física (devido a dificuldades de acionamento) tem consequências sobre sua tarefa, como perda de tempo, perda de dados, retrabalho, repetição da tarefa, etc (CYBIS, 1997).

Segundo Cybis (2007), um problema de usabilidade será hipotético, quando diagnosticado por meio de uma avaliação de ergonomia, ou comprovado, quando observado por um teste de usabilidade. Ou seja, enquanto a usabilidade é *medida*, a ergonomia da interface só pode ser inspecionada e/ou avaliada a partir de recomendações e critérios ergonômicos.

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. Dentre os métodos sem participação de usuários podem ser citados: a avaliação heurística, avaliações analíticas, inspeções baseadas em listas de verificação e inspeções cognitivas. Como métodos que necessitam da presença de usuários para sua realização podem ser citados: a análise da atividade, grupo focal, avaliação cooperativa, testes de usabilidade e questionários de satisfação.

2.1.2.1 Avaliação Heurística

Descreve um método no qual um grupo de avaliadores examina uma determinada interface à procura de problemas que violem um conjunto de heurísticas. É um método simples, com duração de uma a duas horas e de menor custo que outros métodos. A eficiência desse método reside na capacidade dos avaliadores de reconhecerem problemas de usabilidade, sendo que qualquer pessoa pode ser treinada para aplicação deste método, embora resultados melhores sejam obtidos com avaliadores experientes (NIELSEN, 1993).

A avaliação heurística pode ser feita por um único avaliador, porém Nielsen (1993) recomenda que sejam empregados de três a cinco avaliadores. Cada avaliador deve realizar a sua inspeção individualmente e, durante a avaliação, deve ser evitada a comunicação entre eles, garantindo assim, avaliações independentes e sem influência. Durante a realização da avaliação, o avaliador percorre a interface diversas vezes, inspecionando os diversos elementos da interface e comparando-os com a lista de heurísticas de usabilidade.

As avaliações heurísticas, no entanto, podem apresentar inconvenientes, como a possibilidade de produzirem diagnósticos equipados e sugestões de revisão superficiais. Tais

fatos decorrem da falta de conhecimento mais aprofundado tanto sobre o contexto de uso dos sistemas como sobre o contexto do projeto da interface (CYBIS, 2007).

2.1.2.2 Avaliação Analítica

Esta técnica, segundo Cybis (2007), enfoca a estrutura da tarefa com um dispositivo informatizado e se caracterizam por algum tipo de decomposição da tarefa para verificar a complexidade ou estimar os tempos das interações propostas. É empregada nas primeiras etapas da concepção de uma interface.

Deve-se ter em mente que, ao aplicar essa técnica, não se estará levando em conta a mínima hesitação do usuário, erro ou incidente de interação. Trata-se de prever o tempo da interação perfeita. Esse tipo de análise pode ser útil para comparar alternativas de interfaces, mesmo preservando seus limites.

2.1.2.3 Inspeção baseada em listas de verificação (Checklist)

Esta técnica é baseada em um conjunto de questões baseadas em recomendações, que são aplicáveis diretamente na interface por programadores e analistas, não necessariamente especialistas em usabilidade, fazendo rapidamente o diagnóstico em busca de problemas gerais e repetitivos da interface, sem exigir um grande esforço de interpretação (WINCKLER, 2001). Nesse tipo de técnica, ao contrário da avaliação heurística, são as qualidades do *checklist*, e não dos avaliadores, que determinam a qualidade da avaliação.

Para Winckler (2001) a aplicação de *checklists* é um tipo de avaliação rápida e de baixo custo e pode ser adaptada às diversas situações de avaliações, bastando para tanto selecionar as regras ergonômicas adequadas.

Entretanto, segundo Cybis (2007, p.188)

“deve-se ter muita atenção com a qualidade das listas de verificação. Listas mal elaboradas, que apresentam questões subjetivas e número insuficiente, pode levar à produção de resultados duvidosos, pouco uniformes e pouco abrangentes. Os resultados podem ser ainda mais duvidosos caso existam questões inadequadas, que solicitem do inspetor um nível de competência em usabilidade ou de conhecimento sobre o contexto de operação do sistema que ele não possui. Por outro lado, a economia na inspeção fica prejudicada por listas que proponham grande quantidade de questões, que em sua maioria não são aplicáveis ao sistema em avaliação.”

Listas de verificação podem ser desenvolvidas para um contexto, baseadas em recomendações genéricas como as da ISO 9241, ou as fornecidas pelo site ErgoList (<http://www.labiutil.inf.ufsc.br/ergolist>), desenvolvido pelo LabIUtil, Laboratório de Utilizabilidade da Universidade Federal de Santa Catarina.

2.1.2.4 Inspeção Cognitiva

Inspeção Cognitiva é uma técnica de avaliação que verifica os processos cognitivos do usuário enquanto ele interage com o sistema pela primeira vez (KIERAS; POLSON, 1999).

Para aplicar essa técnica, o inspetor deve conhecer o caminho previsto no sistema para a realização das principais tarefas e deve também levar em consideração aquilo que o usuário já conhece sobre a tarefa e sobre a operação de sistemas. De posse dessas informações, ele passa a percorrer os caminhos previstos, aplicando, para cada etapa, a seguinte lista de verificação:

- O usuário tentará realizar a tarefa certa?
- Ele verá o objeto associado a essa tarefa? O objetivo está suficientemente à vista do usuário?
- Ele reconhecerá o objeto como associado à tarefa?
- Ele saberá operar o objeto?
- Ele compreenderá o feedback fornecido pelo sistema como um progresso na tarefa?

A proposta dos autores dessa técnica é que os próprios projetistas possam aplicá-la no desenvolvimento do sistema interativo (CYBIS, 2007).

2.1.2.5 Análise da Atividade

Na análise da atividade, faz-se um levantamento sobre a atividade desempenhada pelo usuário, observando-o em seu ambiente natural, devendo ser realizada antes do início do projeto. Após essa etapa, deve-se criar um protótipo que seja o mais adequado possível a execução das atividades do usuário. Em seguida, deve-se validar o protótipo, apresentando-o ao usuário e através das suas observações, realizar modificações no protótipo original, gerando um novo protótipo. Com a conclusão das alterações solicitadas pelo usuário, realiza-se uma nova apresentação, onde deverá ser revalidado o protótipo, procurando assim, adequá-lo o mais próximo as suas atividades (ARAGÃO, 2001).

2.1.2.6 Grupo Focal

Grupo focal tem sido empregado em pesquisas mercadológicas desde a década de 1950 e a partir da década de 1980 começaram a despertar o interesse dos pesquisadores em outras áreas do conhecimento, como as ciências Sociais, a Ergonomia, as Ciências Médicas, a Ciência da Informação, entre outras (DIAS, 2000).

O principal objetivo desta técnica é identificar percepções subjetivas, grau de satisfação, atitudes e opiniões dos usuários envolvidos. Nesta técnica, aproximadamente seis a nove usuários juntamente com um moderador são reunidos para discutir novos conceitos e identificar questões por um período de aproximadamente duas horas. Para preparar um grupo focal o moderador necessita preparar uma lista de questões para ser discutida, e um conjunto de metas para o tipo de informação que estão sendo reunidas. Durante a seção em grupo, o moderador necessita manter a discussão no foco sem inibir o livre fluxo de idéias e comentários. Entretanto, como todos os métodos que são baseados no questionamento direto aos usuários sobre o que eles querem, em vez de medir ou observar como eles realmente usam a interface, grupos focais envolvem o risco de que os usuários podem pensar que elas querem determinada coisa, porém, eles de fato necessitam de outra. Este problema pode ser reduzido pela exposição ao usuário de um exemplo concreto da tecnologia em discussão (NIELSEN, 1993).

2.1.2.7 Avaliação Cooperativa

Segundo Monk *et al* (1993), a avaliação cooperativa é um procedimento para obter dados sobre problemas experimentados ao se trabalhar com um protótipo de um *software*, de modo que realize mudanças para melhorá-lo. O que distingue este tipo de avaliação é a cooperação que ocorre à medida que usuários e *designers* avaliam o sistema juntos. Os usuários são encorajados a perguntar ao avaliador sobre o processo de interação com o sistema. O avaliador, por sua vez, questiona sobre o entendimento em relação ao sistema.

A avaliação cooperativa funciona da seguinte forma: os usuários trabalham em tarefas representativas escolhidas pelo *designer* e, à medida que trabalham, eles explicam para o *designer* o que estão fazendo e fazem perguntas; o *designer* permite que o usuário cometa erros e utiliza suas perguntas para obter mais informações sobre problemas em potencial,

comportamentos inesperados e comentários a respeito da interface são vistos como sintomas de problemas de usabilidade em potencial.

2.1.2.8 Teste de Usabilidade

Teste de Usabilidade corresponde a observação da interação de usuários no mundo real ou sob condições controladas. Os avaliadores reúnem os dados dos problemas detectados no uso e verificam se a interface suporta o ambiente e as tarefas do usuário (MATIAS, 1995).

Dumas e Redish (1994) definem o teste de usabilidade por meio de cinco características: (i) o objetivo é prover a usabilidade do produto, (ii) os participantes representam usuários reais, (iii) os usuários executam tarefas reais, (iv) os avaliadores observam e registram os participantes e (v) os avaliadores então analisam os dados e recomendam mudanças.

Em um teste de usabilidade, não se trata mais de elaborar diagnósticos de problemas de usabilidade em função de uma desconformidade quanto à ergonomia na interface. O objetivo é constatar esses problemas, medir seu impacto negativo sobre a interação e identificar suas causas na interface. É uma técnica bem mais elaborada, que envolve uma simulação de situações de uso do sistema. E a facilidade ou dificuldade para a realização deste teste dependerá do nível de exigência requerido para os resultados, da generalidade do produto e a disponibilidade de recursos e de usuários (CYBIS, 2007).

Segundo Nielsen (1993), enquanto se prepara um teste de usabilidade, é preciso estabelecer seus objetivos, pois este é um ponto que produz impactos relevantes no tipo de teste a ser realizado, e a principal distinção é se o teste visa obter uma ajuda no desenvolvimento ou se o teste tem por objetivo avaliar a qualidade global da interface. Quando o teste visa somente uma ajuda, seu foco é relatar os aspectos bons e ruins, assim, verificando como o design pode ser melhorado, neste caso se recomenda o teste denominado “Pensar em Voz Alta”, ou seja, “*Thinking-aloud test*”. Enquanto que quando o teste tem por objetivo efetuar uma análise global da interface em fase final de definição, recomendam-se testes que ofereçam medidas de desempenho.

Nielsen (1993) ainda afirma que, independente do tipo de teste a ser aplicado, deve-se desenvolver um detalhado plano de teste, em que se deve levantar as seguintes questões:

- O objetivo do teste: O que se deseja obter?

- Onde e como o teste irá acontecer?
- Qual a previsão de duração de cada sessão?
- Que suporte computacional, qual será necessário?
- Qual software terá que estar disponível para o teste?
- Em que estado deverá estar o sistema para o teste?
- Qual será o tempo de resposta do sistema?
- Quem serão as pessoas que irão experimentar o sistema?
- Quem serão os usuários e de que maneira serão conseguidos?
- Quantos usuários serão necessários?
- Quais tarefas os usuários precisarão executar?
- Qual critério será empregado para definir se o usuário acabou a tarefa de forma correta ou incorreta?
- O usuário terá apoio (manual, ajuda online, etc) durante o teste?
- O experimentador poderá auxiliar os usuários durante o teste?
- Quais serão os dados coletados e de que maneira serão analisados depois de coletados?
- Qual critério será empregado para determinar se a interface é ou não um sucesso?
(p. ex: não há nenhum problema de usabilidade com severidade maior ou igual a 3)

2.1.2.9 Questionários de satisfação

Desde a década de 80, os questionários para avaliação de satisfação do usuário vêm sendo pesquisados. Estes questionários devem ser cuidadosamente elaborados pois através dessa ferramenta é que o avaliador poderá obter e concluir os resultados (CHIN; DIEHL; NORMAN, 1988).

Winckler (2001) afirma que questionários podem ser úteis de diferentes maneiras dentro do desenvolvimento de interfaces Web, como, por exemplo, para:

- identificação do perfil dos usuários - o objetivo é coletar informações de origem funcional, pessoal, sobre as preferências ou sobre a utilização de computadores e sistemas, entre outras;
- determinação do grau de satisfação dos usuários com relação à interface;
- estruturação das informações sobre problemas de usabilidade identificados por usuários.

Entre os questionários de satisfação mais conhecidos estão:

- SUS – *System Usability Scale*
(<http://www.usability.serco.com/trump/documents/Suschapt.doc>);
- SUMI – *Software Usability Measurement Inventory*
(<http://www.ucc.ie/hfrg/questionnaires/sumi/index.html>)
- QUIS - *Questionnaire for User Interaction Satisfaction*
(<http://www.lap.umd.edu/QUIS/index.html>)
- WAMMI - *Web local Analysis and Inventory of Measure*
(<http://www.wammi.com/>)
- ISONORM 9241/110 – Questionário baseado na parte 110 da norma ISO 9241
(<http://www.sozialnetz-hessen.de/ergoonline/Software/Isonorm-Workshop.htm>)

No Quadro 2.2 tem-se um resumo de algumas das técnicas mais usuais de avaliação de usabilidade.

Técnica	Descrição	Vantagens	Desvantagens
Avaliação Heurística	Análise de conformidade do sistema diante de padrões de qualidade ou heurísticas definidas por especialista.	<ul style="list-style-type: none"> - Rapidez de aplicação, - Baixo custo, - abrangência e - facilidade de identificação dos problemas 	<ul style="list-style-type: none"> - 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.	<ul style="list-style-type: none"> - Avaliação barata, rápida, sistemática e de fácil aplicação. - Outra pessoa não especializada em usabilidade pode aplicar. 	<ul style="list-style-type: none"> - 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.	<ul style="list-style-type: none"> - 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 cuidadosa de tarefas. 	<ul style="list-style-type: none"> - 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	<ul style="list-style-type: none"> - Os problemas são vistos do ponto de vista real - Identifica problemas rotineiros de uso 	<ul style="list-style-type: none"> - Requer tempo e pessoas - Pode não ser totalmente imparcial.
Grupo Focal	Discussão entre seis a nove usuários, orientadas por um moderador.	<ul style="list-style-type: none"> - Promove a participação do usuário. - Boa fonte de novas idéias e/ou sugestões 	<ul style="list-style-type: none"> - Requer tempo e pessoas - Subjetividade
Avaliação Cooperativa	Usuários e projetistas discutem juntos a interface.	<ul style="list-style-type: none"> - Fornece idéias para o projeto através das sugestões dos usuários. - Promove a participação do usuário. - Experiência educacional rica. 	<ul style="list-style-type: none"> - Requer tempo e pessoas - Subjetividade
Teste de usabilidade	Simulação de situações de uso objetivando constatar problemas medir seus impactos e identificar suas causas.	<ul style="list-style-type: none"> - Abrangência - Verificação de situações reais de uso 	<ul style="list-style-type: none"> - Alto custo - Requer tempo e pessoas - Pode não ser totalmente imparcial.

Quadro 2.2 - Descrição, vantagens e desvantagens das técnicas de avaliação de ergonomia de interface e usabilidade estudadas.

Todas estas técnicas apresentadas podem ser utilizadas para avaliar as mais diversas interfaces, podendo ser *softwares*, jogos eletrônicos, dispositivos móveis, internet entre outros. Dentro da internet, a interface pode ser pessoal, comunicativa ou comercial, como é o caso do e-commerce, que é o foco do presente trabalho. Para entender a estrutura de usabilidade presente em um site de e-commerce e conseqüentemente, propor uma avaliação deste, é necessário antes conhecer alguns conceitos e peculiaridade deste setor.

2.2 Comércio Eletrônico

2.2.1 Conceito

O comércio eletrônico pode ser definido como a compra e venda de informações, produtos e serviços, além de todos os serviços de pré e pós-venda realizados através de computadores, incluindo-se nesse segmento a Internet e todas as mídias interativas englobando vários processos de negócios da organização (GÓIS, 2000). Outros autores (BLOCH; PIGNEUR; SEGEV, 1996; KALAKOTA; WHINSTON, 1997; CAMERON, 1997; ALBERTIN, 2004) definem comércio eletrônico de forma semelhante, utilizando nomenclaturas um pouco diferente ou diferentes aprofundamento, entretanto, todos argumentam que as operações de compra e venda se dêem por via eletrônica.

Segundo Cullen e Webster (2007), a última definição de *e-commerce* que tem sido adotada no âmbito empresarial é a do *Office for National Statistics (ONS)*, que classifica de *e-commerce* a utilização de redes eletrônicas para realização de pedidos (utilizando todos os tipos de redes mediadas por computadores) e não simplesmente como um canal de pagamentos ou de entrega.

2.2.2 Tipos de comércio eletrônico

Intrinsecamente ao *e-commerce*, várias novas formas de mercado surgiram. Dentre elas, destacam-se:

- *B2B (business to business)* – comércio entre empresas;
- *B2A (business to administration)* – comércio entre empresa e governo;
- *C2C (consumer to consumer)* – comércio entre consumidores;

- *C2A (consumer to administration)* – comércio entre consumidor e governo;
- *B2C (business to consumer)* – comércio entre empresa e consumidor.
- *M-commerce* - quando o comércio eletrônico ocorre em um ambiente de comunicação sem fio, como, por exemplo, utilizando telefones celulares para acesso à Internet.

A Figura 2.2 demonstra algumas possibilidades de comércio.

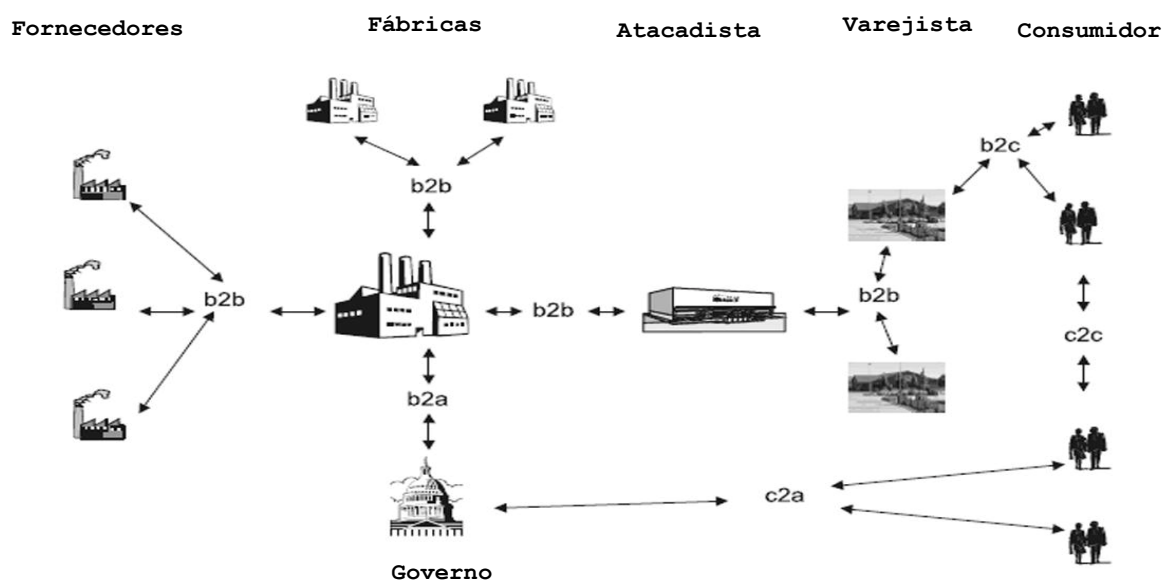


Figura 2.2 - Tipos de Comércio Eletrônico.
Fonte: Adaptado de Delfmann; Albers; Gehring (2002)

O comércio eletrônico B2B (business to business) caracteriza-se por empresas fornecedoras que desenvolvem *sites* na Internet, através dos quais as empresas clientes podem obter e trocar informações com os fornecedores, como também adquirir produtos (NOVAES, 2004).

No sistema B2B, algumas empresas vêm utilizando a Internet no contato entre fornecedores e clientes para agilizar suas transações comerciais. Outras têm se valido de portais setoriais, mercados virtuais de compra, os e-marketplaces (mercados na internet, também chamados de comunidades), e outras têm optado pela construção de nets próprias e dedicadas (BIO et al., 2001).

No B2C (business to consumer), o comprador é uma pessoa física que, a partir de um computador pessoal, realiza suas buscas e adquire um produto ou serviço através da Internet

(NOVAES, 2004). Ele refere-se à realização de negócios pela Internet, entre as empresas e os consumidores finais, apresentando muitos aspectos singulares em relação ao varejo tradicional de produtos físicos.

Apesar das empresas e comércio estarem se adaptando a nova realidade, o que significa que muitas ineficiências ainda ocorrem no processo, o número de clientes no mercado B2C tem aumentado a cada ano.

De acordo com as pesquisas realizadas pela empresa e-bit (2009), o número de consumidores de e-commerce em 2008 cresceu 39% em relação a 2007, o que totaliza um mercado de cerca de 13,2 milhões de consumidores. Neste mesmo período as vendas cresceram 30% o que representa um faturamento de R\$ 8,2 bilhões em vendas de produtos pela Internet no Brasil apenas no ano de 2008. Ainda segundo a pesquisa, o valor médio de compra de um usuário em uma compra *on-line* foi de R\$ 328.

2.2.3 O consumidor no comércio eletrônico

Gabbott e Hogg (1998) argumentam que a interação entre consumidor e vendedor no *e-commerce* segue modelos padrões representados por diferentes estágios de interação.

Segundo Minocha *et al* (2006), a literatura acerca do comportamento do consumidor em *e-commerce* oferece modelos padrões de interação entre consumidor e vendedor, que na maioria dos casos é representada por diferentes estágios de interação. Como o proposto por Gabbott e Hogg (1998) onde a interação envolve três estágios:

- O estágio de pré-compra, durante a qual o consumidor passa pelo processo de encontrar o produto ou serviço, esta fase evidencia as esferas de marketing, acessibilidade e usabilidade do meio de interação;
- O estágio de compra em si, onde o consumidor realiza a compra, neste estágio, além das características de usabilidade são priorizados também questões de privacidade e segurança;
- O estágio de pós-venda, que engloba, entrega de produtos ou serviço, suporte técnico etc.

Petre, Minocha e Roberts (2006) argumentam que durante a interação, a avaliação do consumidor envolve dois aspectos:

- A usabilidade do site de *e-commerce*, que é determinada pela interface gráfica, pelo design de interação, elementos visuais, navegação, arquitetura da informação, e tantos outros
- A qualidade do serviço, que está relacionado a expectativa durante os diferentes estágios de interação, assim como aspectos do produto, varejista ou prestador de serviço, confiabilidade, receptividade, empatia, qualidade da informação, satisfação, privacidade e assim por diante.

Experiências insatisfatórias em qualquer um dos estágios de interação podem render uma impressão negativa.

A usabilidade e a qualidade do serviço no ambiente de *e-commerce* determinam o valor percebido pelo consumidor durante o processo de compra. O valor percebido é a avaliação geral do consumidor sobre o serviço e é baseada na percepção do que é recebido e o que é fornecido (alguns consumidores estão preocupados com o dinheiro gasto, outros com o tempo gasto ou o esforço) (SHAW; IVENS, 2002).

Segundo Minocha *et al* (2006) as empresas estão reconhecendo que fornecer valor através da satisfação traz diferencial competitivo a organização. Atendimentos às percepções de valor, às expectativas de usabilidade e qualidade de serviço estão de certa forma condicionadas a alguns fatores de ordem comportamental dos consumidores como:

- fatores individuais, tais como valores pessoais do cliente, as atitudes em relação ao *e-commerce*, opções de compra off-line e experiências anteriores com o *e-commerce*;
- fatores sociais, como as recomendações, tendências de mercado, publicidade, e
- fatores organizacionais, como a fidedignidade do varejista e de experiências anteriores com canais off-line da empresa.

Do ponto de vista cognitivo, Borges; Morales; Rodríguez (1996) argumentam que no desenvolvimento de páginas para *e-commerce* deve-se facilitar o desenvolvimento de modelos mentais, de forma intuitiva, e não impor ao usuário rigor e degraus a serem vencidos para que obtenha as informações que deseja e possa efetivar sua compra.

Um site de *e-commerce* deve eliminar elementos de design confusos e utilizar o máximo possível as convenções de design. Os padrões aprimoram o sentido do domínio do usuário em relação a um site, ajudando-o a realizar suas tarefas e aumentar sua satisfação geral com um site (NIELSEN; LORANGER, 2006). O que vem de encontro ao que Davis (1989) classificou como necessidade de evitar a superposição dos modelos mentais envolvidos.

A sistematização destes conceitos teóricos de usabilidade em e-commerce pode ser realizada de várias maneiras, uma delas pode ser a criação de escalas de medidas, que tenham a capacidade de comparabilidade e representatividade para facilitar a avaliação.

2.3 Escalas de Medidas

Escalas são instrumentos que objetivam medir traços latentes, intensidade de opiniões ou atitudes da maneira mais objetiva possível. Uma escala consiste em um arranjo, em forma de série graduada de itens, pelo qual se mede uma característica, de acordo com um número previamente determinado, e são compostas por categorias, que são as gradações ou alternativas de respostas oferecidas. O que se espera de uma escala é a discriminação de medidas. (GIL, 1999; MOTTA, 1999).

A maior implicação das escalas é a possibilidade de comparação entre escores e apropriadas estatísticas que resumem estes escores. Em uma escala de medida, o papel dos números é de representar um objeto, ou uma pessoa, de forma que a relação entre os números represente a relação empírica entre os objetos. Esta representatividade fornece ao objeto importantes propriedades tais como, poder de distinção entre objetos, ordem, adição e razão (EMBRETSON; REISE, 2000).

2.3.1 Tipologias de escalas

A distinção entre tipos de escala pode ser realizada através de quatro níveis de medidas: escala nominal, escala ordinal, escala intervalar, escala de razão (STEVENS, 1946; EMBRETSON; REISE, 2000; GÜNTHER, 2003; PASQUALI, 1997; SOMMER; SOMMER, 1997).

Na escala nominal, números ou símbolos são utilizados para representar categoria de objetos ou pessoas. Números específicos e símbolos são simplesmente rótulos, propriedades que definem ordem adição ou razão não são atingidas em medições nominais (EMBRETSON; REISE, 2000).

Numa escala ordinal, além de identificarem pessoas, objetos ou categorias, números ou símbolos os ordenam numa dimensão subjacente. Um exemplo da aplicação desta escala é a hierarquização de preferências ou importâncias entre pessoas ou objetos, status social

propriedades de um sistema etc (GÜNTHER, 2003). Dados de uma escala ordinal, satisfazem todas as propriedades de dados nominais e adicionam a representação numérica uma ordem as pessoas ou ao traço latente. Permite transformações e preservam tanto a distinção dos dados quanto a ordem (EMBRETSON; REISE, 2000).

Na escala intervalar, as características não somente podem ser ordenadas conforme uma dimensão subjacente, mas os intervalos entre as alternativas têm tamanho conhecido e podem ser comparados (GÜNTHER, 2003). Nesta escala são permitidas apenas transformações lineares, ou seja, escores podem ser multiplicados por uma constante e ou ter uma constante adicionada (por exemplo, $aX_1 + b$). Parâmetros estatísticos são apropriados para escalas intervalares, tais como média, desvio padrão e correlações (EMBRETSON; REISE, 2000).

Em uma escala de razão, a razão dos escores é significativa (por exemplo, um escore de 6 representa o dobro do escore 3 em determinada propriedade). A única transformação permitida neste tipo de escala é a multiplicação por uma constante que preserve a razão entre os escores (EMBRETSON; REISE, 2000).

A escala *Likert* é mais utilizada nas ciências sociais, especialmente em levantamento de atitudes, opiniões e avaliações. Nela pede-se ao respondente que avalie um fenômeno numa escala de, geralmente, cinco pontos: aplica-se totalmente, aplica-se, nem sim nem não, não se aplica, definitivamente não se aplica. As afirmações podem ser auto-referentes : “Eu considero importante em um site ter acesso em outras línguas além do português”. Ou hétero-referentes: “É importante para a comunidade estrangeira ter acesso em outras línguas além do português”. Dependendo do tema subjacente, as alternativas podem, além da dimensão “aplica-se”, seguir dimensões como: “bom – ruim” ou concordo – discordo” (GÜNTHER, 2003). Gil (1999) acrescenta que a escala de *Likert* é de elaboração simples e tem caráter ordinal ou intervalar, e não mede quanto uma atitude é mais ou menos favorável. É uma escala onde os participantes registram sua concordância ou discordância com um enunciado.

A diferenciação entre os quatro tipos de escala tem consequências importantes quanto a complexidade da análise estatística desejável. Dados obtidos em qualquer escala podem ser apresentados por meio de estatísticas descritivas. Para utilizar estatísticas inferenciais, que permitem ao pesquisador verificar até que ponto determinadas relações ou diferenças são sistemáticas ou não, há de se observar que dados baseados em escalas nominais e ordinais podem ser trabalhados com testes estatísticos paramétricos e não-paramétricos. Dados provenientes de escalas intervalares e de razão vão permitir, além de estatísticas não-paramétricas, procedimentos paramétricos (GÜNTHER, 2003).

2.3.2 Vantagens e desvantagens das escalas

A maior vantagem de uso de escalas é que são instrumentos estruturados e padronizados. Isso permite que a avaliação seja facilmente comparada e contrastada, mesmo para grande número de respostas. As escalas geralmente são fáceis de usar e de entender, o que justifica a ampla aceitação dessa abordagem. Em contraponto, as escalas também podem apresentar algumas desvantagens, como a possibilidade de ocorrerem erros de percepção e de significado percebido. Da mesma forma, a terminologia utilizada na construção da escala pode também influenciar a percepção do respondente, pois termos como “excede as expectativas”, ou “capacidade abaixo da média” podem significar coisas diferentes para pessoas diferentes (SANTOS, 2006).

Outra desvantagem é a possibilidade de ocorrer um erro de avaliação, sendo a tendência central o mais comum. Respondentes ocupados podem tender a responder rapidamente e marcar as opções centrais de uma escala, como “satisfatório”, “adequado”, demonstrando uma atitude excessivamente passiva frente às questões colocadas. Ao se elaborar uma escala, alguns problemas básicos se apresentam: definição de um contínuo; confiabilidade; validade; ponderação e natureza de itens. (Gil, 1999; O'Connor, 2007)

A confiabilidade de uma escala se reflete nos resultados semelhantes e consistentes obtidos após a aplicação repetida com a mesma amostra. Para se verificar a confiabilidade de uma escala, podem ser aplicadas técnicas como a de teste-e-reteste, onde a escala é aplicada duas vezes à mesma população e se comparam os resultados. Alguns pesquisadores denominam esse procedimento de pré-teste. Neste método, a correlação entre as provas é um índice de confiabilidade ou de consistência. A validade de uma escala é observada quando ela mede exatamente aquilo a que se propõe medir. A fim de se verificar o grau de validade de uma escala, os itens podem ser submetidos ao julgamento de um grupo de pessoas consideradas especialistas em determinado domínio do conhecimento. Outra forma de verificar a validade de uma escala é submetê-la ao crivo de grupos conhecidos distintos, como votantes de esquerda e de direita. Se forem observadas distinções entre os resultados dos dois grupos, então a escala tem validade (SANTOS, 2006).

Outra fonte de problema em potencial está relacionada à ponderação dos itens. As formas utilizadas para se verificar a validade também podem ser aplicadas para se verificar a ponderação dos itens. Gil (1999) destaca, também, o problema relacionado à natureza dos itens, que podem ser diretos ou projetivos. Os itens diretos apelam para respostas diretas de

concordância ou discordância, ao passo que os projetivos são apresentados como estímulos para as respostas, como frases incompletas, por exemplo.

2.3.3 Elaboração de um conjunto de itens

Conjunto de itens, ou questionários, é composto por uma quantidade razoavelmente elevada de questões apresentadas às pessoas, e uma forma de registrar suas respostas (GIL, 1999; MOTTA 1999). O objetivo é conhecer as opiniões e atitudes, de maneira objetiva, em relação à interface em questão. Os questionários podem ser compostos, de maneira geral, por perguntas fechadas ou perguntas abertas. A existência de cada tipo de pergunta dependerá do tipo de dado a ser coletado. Em questões fechadas, o usuário especifica se concorda, se discorda ou se está indeciso, também especifica sua preferência ou concordância, e, ainda, pode escolher um item em uma lista. As perguntas abertas podem resultar em boas idéias, mas são mais difíceis de analisar e de quantificar. Sempre é necessário realizar um teste-piloto com o questionário, afim de validá-lo. Um questionário para coletar dados referentes à satisfação do usuário precisa ser elaborado de maneira a possuir bom nível de confiabilidade (poder ser aplicado em situações similares e obter os mesmos resultados) e bom nível de validade (saber se são obtidos exatamente os dados necessários) (SANTOS, 2006).

A elaboração de um questionário, segundo Schuman e Kalton (1985) parte da definição do(s) objetivo(s) da pesquisa, que no presente trabalho é medir a usabilidade em *sites* de *e-commerce*. Este(s) objetivo(s) segundo os autores levam necessariamente em consideração a relação conceito/item e a relação população-alvo/amostra, como ilustrado na Figura 2.3.

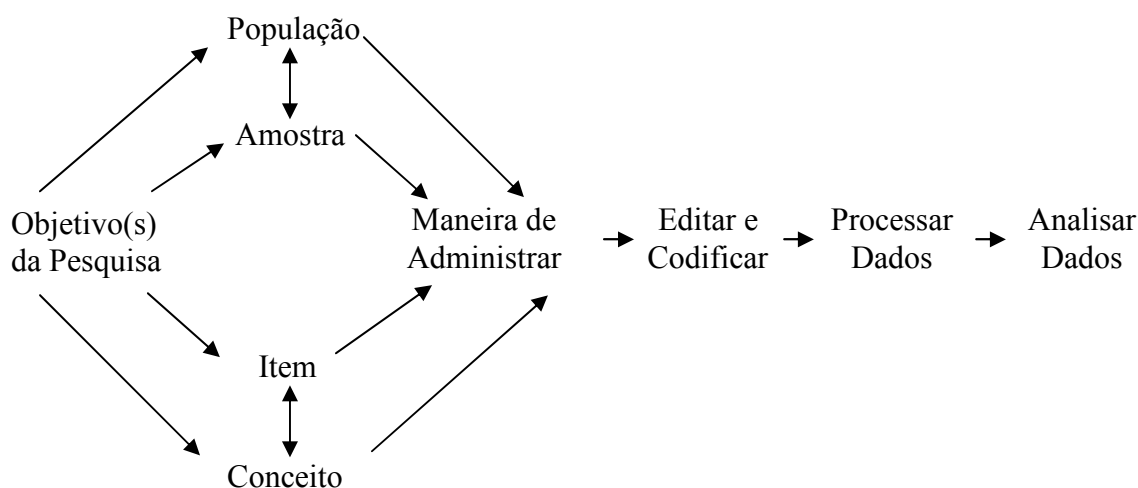


Figura 2.3 - Principais estágios de um survey
Fonte: Adaptado de Schuman e Kalton (1985)

Conceito e população representam os termos abstratos da pesquisa, traduzidos em termos práticos pelo item e pela amostra.

O modelo proposto por Schuman e Kalton (1985) indica também uma relação direta entre população-alvo e amostra assim como entre conceito e item, atrelados ao objetivo da pesquisa. Desta forma, dependendo do objetivo da pesquisa e da população-alvo, definir-se-ão diferentes tipos de amostra, o mesmo ocorre entre conceito e item.

Na sequência os conceitos, o conteúdo dos itens, a população-alvo e a amostra determinam o instrumento e a maneira de administrar. Segundo Günther (2004) a interdependência entre a elaboração de um instrumento e a estratégia de sua aplicação está nos seguintes pontos:

- (1) o grau de complexidade dos conceitos determina o número de itens e a forma de apresentação deles;
- (2) existe relação recíproca entre as características da população-alvo e a complexidade dos conceitos a serem investigados. Ambos determinam a maneira de transformação dos conceitos em itens e sua aplicação;
- (3) o tamanho da amostra influencia a maneira de aplicar o instrumento, em termos de entrevista *versus* questionário e em termos de tamanho.

A utilização de item para avaliar usabilidade é uma técnica de análise comumente utilizada na literatura e no meio empresarial, também conhecida como inspeções de usabilidade por *checklists*. Segundo Jeffries *et al* (1991) são vistorias baseadas em listas de verificação, através das quais profissionais, não necessariamente especialistas em ergonomia, como por exemplo, programadores e analistas, diagnosticam rapidamente problemas gerais e repetitivos das interfaces.

Segundo Cibys (2000), neste tipo de técnica, ao contrário das avaliações heurísticas, são as qualidades da ferramenta (*checklists*) e não dos avaliadores, que determinam as possibilidades para a avaliação. *Checklists* bem elaborados devem produzir resultados mais uniformes e abrangentes, em termos de identificação de problemas de usabilidade.

Os resultados obtidos, através dessa técnica, dependem então da organização e do conteúdo geral ou específico, dessas ferramentas. Versões especializadas de um *checklist* podem ser desenvolvidas a partir de um outro, com questões genéricas. As questões do *checklist* podem vir acompanhadas de notas explicativas, exemplos, e de um glossário, a fim de esclarecer possíveis dúvidas associadas às mesmas. Entretanto, muitas vezes, a sistematização é prejudicada devido a questões subjetivas, que solicitam do inspetor um nível de competência

em usabilidade ou de conhecimento sobre o contexto que ele não possui. Outras vezes a abrangência das inspeções é prejudicada devido ao conteúdo incompleto e organização deficiente das listas.

Jeffries *et al.* (1991) mostram que este tipo de técnica proporciona a identificação de uma grande quantidade de pequenos problemas de usabilidade que se repetem nas interfaces dos sistemas.

Baker (2001) afirma que cada item de um conjunto de itens deve medir alguma faceta da habilidade particular de interesse, e do ponto de vista puramente técnico, cada item deve ser de resposta livre, para que o examinado possa escrever qualquer resposta que lhe possa ser apropriada. A construção dos itens para este trabalho partiu de um levantamento bibliográfico da literatura sobre usabilidade em *e-commerce* e buscou ser o mais objetivo possível a fim de gerar respostas dicotomizadas, para facilitar o tratamento estatístico, que foi realizado com base na Teoria da Resposta a Item.

2.4 Teoria da Resposta ao Item

2.4.1 Conceitos básicos

A utilização de testes ou questionários para medir traço latente é verificada na literatura desde o início do século XX, voltada principalmente para a área de psicologia, na tentativa de criar escalas que pudessem mensurar a inteligência. Em 1905 Binet e Simon elaboraram a Escala Métrica de Inteligência, onde eram utilizados testes que incluíam questões e exercícios sobre figuras, números, letras e palavras, em ordem crescente de dificuldade, e eram adaptados às diferentes idades (MADER, 2003). Nesta escala, o resultado indicava a idade mental, que se fosse igual à idade cronológica, significava inteligência normal. Esta escala foi aprimorada mais tarde por Terman (1916).

A validação dos instrumentos de medida, no caso questionários, ainda era muito questionada, como se pode verificar em Thurstone (1928 p.547), o qual afirma que “um instrumento de medida, na sua função de medir, não pode ser seriamente afetado pelo objeto de medida. Na extensão em que sua função de medir for assim afetada, a validade do instrumento é prejudicada ou limitada.” Na tentativa de aperfeiçoar estes testes, Thurstone (1927) utilizou o modelo de distribuição normal em sua escala de dispersão, o que foi

segundo Van Der Linden e Hambleton (1997) a base para o desenvolvimento do modelo da ogiva normal. Este modelo foi aperfeiçoado por Richardson (1936), Mosier (1941, 1942), Ferguson (1942) e Lawley (1943), que forneceram alguns métodos para estimar os parâmetros dos itens, afastando-se teoricamente da teoria clássica dos testes. Nesta época surgem também os conceitos curva característica do item (TUCKER, 1946) e do traço latente (LAZARSELD, 1950), que configuram conceitos chave na Teoria da Resposta ao Item (TRI).

Segundo Van Der Linden e Hambleton (1997), o modelo da ogiva normal foi o primeiro modelo da TRI, pois postulou uma função normal de distribuição cumulativa como uma função resposta para o item, apresentada na Equação (2.1), onde b_1 representa o parâmetro de dificuldade e θ é o ponto na escala de habilidade onde o indivíduo tem 0,5 de probabilidade de acerto do item, e a_1 é proporcional à inclinação da tangente da função resposta neste ponto, e z representa o desvio padrão.

$$P_i(\theta) = \int_{-\infty}^{a_1(\theta-b_1)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \quad (2.1)$$

No entanto, apesar de já conceituar os parâmetros de dificuldade e de discriminação do item que se tem atualmente, esta abordagem apresentava algumas incoerências conceituais e matemáticas, que começaram a ser equacionadas por Lord (1952), que desenvolve o modelo unidimensional de dois parâmetros baseado no modelo da ogiva normal.

Birnbaum (1968), baseado no trabalho de Lord (1952), sugere a troca da função ogiva normal pelo modelo logístico de dois parâmetros (Equação 2.2), com a finalidade de facilitar o tratamento matemático. 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.

$$P(U_{ij} = 1 / \theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (2.2)$$

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 . E D é um fator de escala constante e igual a 1. Utiliza-se o valor 1,7 quando se deseja que a função logística forneça resultado semelhante ao da função ogiva normal. Segundo Andrade, Tavares e Valle (2000 p.11) “O parâmetro a é proporcional à derivada da tangente da curva no ponto de inflexão.” E quanto maior for o valor do parâmetro a_i , mais sensível torna-se o modelo a variações na habilidade em relação ao seu ponto de dificuldade. Ou seja, valores altos de a_i resultaram em uma maior distinção entre respondentes com traço latente diferente em torno da escala de dificuldade.

Birnbaum (1968) introduziu o terceiro parâmetro modelando um acerto ao acaso, representado pelo parâmetro c_i na Equação (2.3). Infelizmente, ao fazê-lo, algumas das propriedades matemáticas da função logística foram perdidos. O modelo resultante tornou-se conhecido como o modelo logístico de três parâmetros, ainda que tecnicamente não seja mais um modelo logístico (BARKER, 2001).

$$P(U_{ij} = 1 / \theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}} \quad (2.3)$$

Barker (2001) argumenta que, por definição, o valor de c não varia em função do nível da habilidade. Assim, os respondentes com a mais baixa e mais alta capacidade têm a mesma probabilidade de acerto casual. Segundo o mesmo autor o parâmetro c tem um intervalo teórico de $0 \leq c \leq 1,0$, mas, na prática, valores acima de 0,35 não são considerados aceitáveis.

A Figura 2.4 mostra a curva característica do item (CCI) de dois itens hipotéticos e a identificação dos parâmetros dos itens. O eixo x representa a dificuldade dos itens e está na mesma escala do grau de usabilidade, ou seja, quanto maior a dificuldade maior o grau de usabilidade requerido com aquele item. Sendo assim, o parâmetro b que representa matematicamente a dificuldade do item i é definido como o valor onde a probabilidade de resposta positiva ao item é de 0,5. O parâmetro a_i que representa a discriminação do item é proporcional à derivada da tangente da curva no ponto de inflexão e indica a inclinação da curva, ou seja, quanto maior for o valor de a_i , mais inclinada vai ser a curva e, consequentemente mais estreito será o intervalo de discriminação do item i .

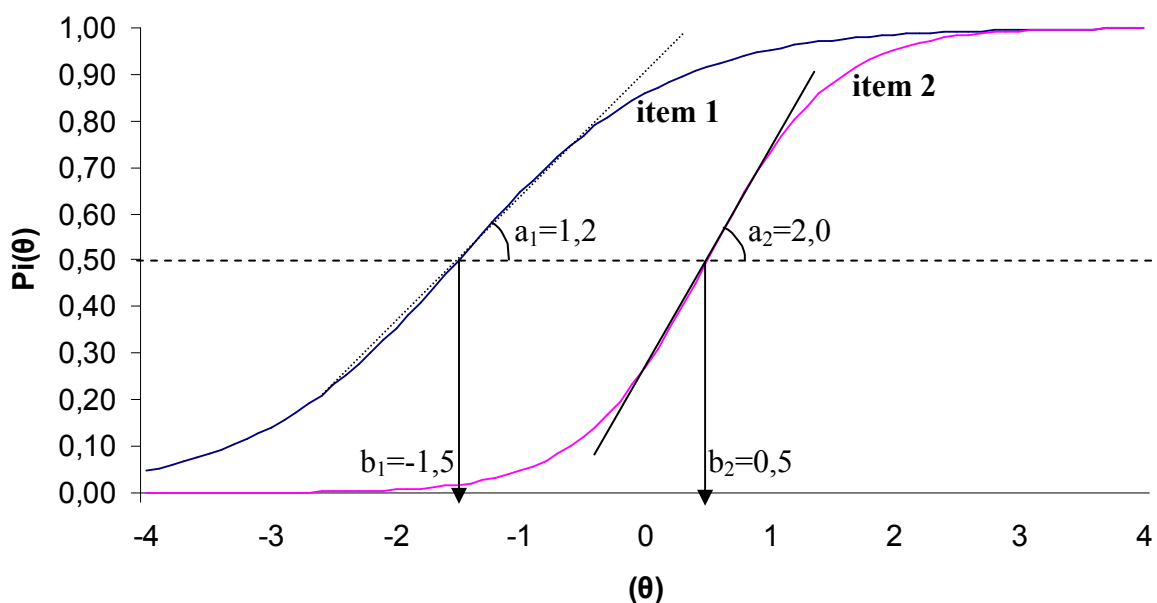


Figura 2.4 - Curva Característica do Item (CCI)
Fonte: Adaptado de Pasquali e Primi (2003)

A escala gerada é arbitrária, e medida na mesma unidade que o parâmetro b , e o que interessa neste caso são as relações ordinais existentes e não necessariamente sua magnitude. Normalmente é pré-suposto que a escala assuma uma métrica normal (0, 1) ou seja, media 0 (zero) e desvio padrão 1 (um).

Birnbaum (1968) introduziu a medida Fisher, que descreve a estrutura de informação do teste, representada pela soma de todas as informações individuais de cada item. Segundo Pasquali (2003), a função de informação do item é um poderoso método para descrever itens, bem como para selecionar itens, pois permite analisar quanto um item traz de informação para a medida em questão. A fórmula da função de informação de item é representada pela Equação (2.4).

$$I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad (2.4)$$

Onde,

$I_i(\theta)$ é a informação fornecida pelo item i ao nível da aptidão θ ;

$P_i'(\theta)$ é a derivada de $P_i(\theta)$ em relação a θ ;

$P_i(\theta)$ é a curva característica do item e

$Q_i(\theta) = 1 - P_i(\theta)$.

Paralelamente ao trabalho de Birnbaum e Lord nos EUA, o dinamarquês Rasch (1960) que nos anos 40 já trabalhava com medidas em educação e em psicologia, desenvolveu, baseado na distribuição de Poisson e no modelo de ogiva normal, um modelo unidimensional diconômico que pode ser verificado na Equação (2.5) (Van Der LINDER; HAMBLETON 1997).

$$P_i(U_{ij} = 1 | \theta_j) = \frac{\theta_j}{\theta_j + \delta_i} \quad (2.5)$$

Onde θ_j representa a habilidade do examinado j e δ_i representa a dificuldade do item i . Os estudos de Rasch (1960) influenciaram Andersen (1973) que trabalhou métodos de estimação estendendo o modelo de Rasch a dados binários. No entanto foi a visita de Rasch aos EUA que inspirou estudiosos como Wright e Douglas (1977) a desenvolver uma modelagem logística ao modelo original de Rasch (1960) (EMBRETSON e REISE 2000).

Como afirmam Van Der Linden e Hambleton (1997) formalmente, os modelos de Rasch (1960) são um caso particular dos de Birnbaum (1968), mas que conduzem a uma teoria da análise do teste completamente diferente.

Lord e Novick (1968) publicam o livro intitulado *Statistical Theories of Mental Test Score* onde fornecem um rigoroso e unificado tratamento estatístico incluindo os conceitos da TRI até então desenvolvidos e comparando com a Teoria Clássica dos Testes.

No entanto, os modelos até então desenvolvidos consideravam apenas respostas dicotômicas. Samegima (1969) elaborou modelos para tratar respostas politômicas ou graduais objetivando assim obter respostas mais completas do respondente. Os estudos de Samegima (1969) também receberam contribuições de Bock (1972), Andrich (1978), Masters (1982) e Muraki (1992).

Todavia, um ponto crítico na TRI era a estimação dos parâmetros envolvidos nos modelos, em particular quando se necessita estimar tanto os parâmetros dos itens quanto as habilidades (ANDRADE; TAVARES; VALLE, 2000). Esta dificuldade foi de certa forma facilitada pelo desenvolvimento, com auxílio de ferramentas computacionais, do método da máxima verossimilhança marginal (BOCK; AITKEN, 1981). A partir daí, Bock e vários pesquisadores desenvolveram softwares como BILOG, TESTFACT, MULTILOG e PARSCALE, utilizando este método de estimação (EMBRETSON; REISE 2000).

A introdução de modelos logísticos de 1, 2 e 3 parâmetros para duas ou mais populações de respondentes, foi feita por Bock e Zimowski (1997) possibilitando comparações de rendimentos de duas ou mais populações submetidas a diferentes testes com itens comuns (HEDGES; VEVEA 1997).

Sendo assim, de uma forma resumida, a Teoria da Resposta ao Item (TRI) é um conjunto de modelos 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.

Atualmente, além da área educacional e psicológica, a TRI vem sendo aplicada e explorada em outras áreas, tais como medicina (DeROOS; MEARES, 1998; VIDOTTO ET AL., 2005; DAS e HAMMER, 2005); psicossocial (GRANGER; DEUTSCH, 1998); marketing (BAYLEY, 2001; SINGH, 2004); área de serviços (COSTA, 2001); gestão pela qualidade total (ALEXANDRE ET AL., 2001); sistemas de informação (WU, 1999); genética (TAVARES; ANDRADE; PEREIRA, 2004); psiquiatria (CURI, 2006); qualidade de vida (GUEWEHR, 2007; METZ ET AL, 2006; CELLA, CHANG, 2000); capital intelectual (VARGAS, 2007).

2.4.2 Questões críticas para a aplicação da Teoria de Resposta ao Item

A aplicação da TRI requer algumas pressuposições iniciais tais como a determinação do modelo da TRI que melhor se encaixa ao conjunto de dados a serem analisados e aos objetivos da investigação. Outro pressuposto importante, do qual também depende a determinação do modelo, é a determinação da dimensionalidade do traço latente, ou seja, a quantidade de traços latentes que serão medidos. Segundo Singh (2004) a dimensionalidade refere-se ao número de fatores necessários para explicar a variabilidade dos dados e constituir uma hipótese a ser verificada. A dimensionalidade pode assumir um caráter unidimensional, quando existir apenas um fator em análise, por exemplo, a usabilidade de um site de *e-commerce* ou multidimensional, quando existirem mais de um fator. De um modo geral, segundo Schmitt (1996), quanto mais estritamente unidimensional for o construto, menos

ambígua as interpretações dos resultados e mais legítimas são as correlações. Entretanto, embora a suposição de unidimensionalidade destina-se a medir apenas um traço latente, geralmente a natureza dos testes estatísticos introduz muitas vezes elementos multidimensionais. Quando existir dúvida quanto à unidimensionalidade de um construto, segundo Edelen e Reeve (2007), é importante verificar os resultados do modelo da TRI adotado, a fim de identificar eventuais anomalias que possam surgir devido à violação deste pressuposto, como, por exemplo, um ou mais itens com parâmetro de discriminação com baixíssima declividade pode indicar problema de dimensionalidade.

Portanto, dimensionalidade é um fator intrínseco ao construto interno e define a homogeneidade ou não do conjunto de itens e a desconsideração deste fator resulta, segundo Spencer (2004), em um modelo de medição aplicado indevidamente e gerando inferências errôneas sobre a avaliação dos resultados e pode ameaçar a credibilidade do instrumento.

A literatura traz alguns testes de unidimensionalidade. Reckase (1979) sugere a utilização de uma parcela dos autovalores da matriz de correlação tetracórica. Outra técnica bastante utilizada é a análise fatorial, que é, segundo Spencer (2004) um conjunto de ferramentas estatísticas utilizadas para reduzir o número de variáveis, assim como para avaliar a estrutura de dados, e pode ser exploratória ou confirmatória. No entanto como argumentam Steinberg; Thissen e Wainer (2000) e Pasquali (2003), não existe um consenso em relação ao mais apropriado, cada um tem suas vantagens e desvantagens. Embretson e Reise (2000) argumentam que, infelizmente, a análise fatorial linear exploratória e confirmatória não são apropriadas para dados obtidos de itens respondidos de modos dicotômicos ou politômicos. Em contraponto a isso os autores mencionam que alguns pesquisadores estão utilizando modelos de análise fatorial não-linear ou outras técnicas especialmente desenvolvidas para a análise de itens dicotômicos ou politômicos, dos quais eles sugerem o modelo de análise fatorial de informação plena (full-information), disponível no software TESTFACT (MURAKI; ENGELHARD, 1985; WILSON ET AL, 1991), que é embasada em toda informação disponível na matriz inteira de item-resposta, em contraste a somente a informação da matriz de covariância ou de correlação. Todavia, para Pasquali (2003), se a relação for não-linear, tanto a análise fatorial quanto a análise dos componentes principais produzem resultados que não podem ser interpretados ou que são simplesmente errôneos.

Sendo assim, no presente trabalho, utiliza-se, para verificação de unidimensionalidade os próprios resultados da Teoria da Resposta ao Item. Segundo Pasquali (2003) através dos índices (parâmetros) da TRI é possível analisar a adequação dos seus modelos de análise de

itens. Se estes índices indicarem que o modelo utilizado for adequado, então fica demonstrado que o item é unidimensional. O problema contido neste modo de pensar consiste em que a análise da adequação do modelo deve ser feita depois que foi verificada a existência de unidimensionalidade da série de itens analisados.

A independência local também é um pré requisito básico para a aplicação da TRI, ou seja, os itens devem ser independentes entre si e a probabilidade de responder um item é estritamente determinada pelo nível do traço latente do respondente e não por suas respostas a outros itens do conjunto (ANDRADE; TAVARES; VALLE, 2000; EMBRETSON; REISE, 2000; PASQUALI, 2003).

2.4.3 A escolha do modelo mais adequado

Há registros na literatura de vários modelos paramétricos tanto unidimensionais quando multidimensionais (EMBRETSON; REISE, 2000), no entanto os primeiros mostram-se mais numerosos (Van Der LINDER; HAMBLETON, 1997; EMBRETSON; REISE, 2000; PASQUALI, 2003).

Por se tratar, no presente estudo, da avaliação de apenas um traço latente – a usabilidade em *sites de e-commerce* – considerar-se-ão apenas os modelos paramétricos unidimensionais.

O primeiro ponto a se considerar na escolha de um modelo é o número de categoria de resposta dos itens, dicotômicos (com apenas duas categorias) ou politômicos (com mais de duas categorias).

Para itens dicotômicos os modelos de 1, 2 e 3 parâmetros logísticos são os mais comuns e os modelos de parâmetros de assíntota superior (para 4 parâmetros) também são possíveis (EDELLEN; REEVE, 2007).

Vários critérios podem ser utilizados para determinar qual o melhor modelo, dentre eles (a) os pesos dos itens para pontuação (iguais ou diferentes); (b) as propriedades de escala desejadas para a medição; (c) aderência (ajuste) aos dados e (d) o propósito para estimar os parâmetros (EMBRETSON; REISE, 2000).

Para itens politômicos, os modelos mais comuns são:

- O modelo de Resposta Nominal apresentado por Bock (1972) que utiliza toda a informação contida nas respostas para maximizar a precisão da estimativa do traço latente;

- O modelo de Resposta Gradual de Samejima (1969) que assume que as categorias de resposta de um item podem ser ordenadas entre si e a partir daí obter-se mais informações das respostas;
- O modelo de Escala Gradual formulado por Andrich (1978) que acrescenta ao modelo de resposta gradual a suposição de que os escores das categorias sejam igualmente espaçados, como na escala Likert, fazendo com que o número de categorias de resposta seja o mesmo para todos os itens do conjunto;
- O modelo de Crédito Parcial de Masters (1982) que é uma extensão do modelo logístico de um parâmetro para itens politômicos onde o número de categorias das opções de respostas pode variar de item a item no teste.
- Bock e Zimowski (1997) generalizam os modelos logísticos unidimensionais de 1, 2 e 3 parâmetros para permitir comparação de indivíduos de grupos distintos, submetidos a conjunto de itens diferentes, segundo os autores isto é possível desde que haja um percentual de itens comuns que façam a ligação entre os conjuntos.

2.4.4 Tamanho da amostra

Ainda não existem respostas definitivas quanto aos requisitos do tamanho da amostra. Há algumas orientações gerais, tais como: a necessidade de aumento no número amostral deve ser proporcional ao aumento da complexidade do modelo, a obtenção de menores erros padrões requer maior número amostral, no entanto, pequenas amostras podem ser adequadas para avaliar as propriedades de um conjunto de itens e facilitar ajustes necessários (EDELEN; REEVE 2007).

Linacre (1994) defende que uma amostra em torno de 100 respondentes, é muitas vezes suficiente para estimar de forma estável o modelo logístico de um parâmetro. Tsutakawa e Johnson (1990) recomendam uma amostra de aproximadamente 500 respondentes para estimativas precisas dos parâmetros. No entanto, outros têm sugerido que algo em torno de 200 ou menos observações podem ser adequados (ORLANDO; MARSHALL, 2002; THISSEN, STEINBERG; GERRARD, 1986).

2.4.5 Estimação

Como visto anteriormente, nos modelos da TRI, a probabilidade de uma resposta depende, da habilidade do respondente θ_j , e dos parâmetros que caracterizam o item. Entretanto, nem sempre estes dois fatores são conhecidos, e o que se conhece na maioria dos casos são apenas as respostas do respondente aos itens. Portanto, é necessário se fazer uma estimativa destes parâmetros. Andrade, Tavares e Valle (2000) dividem do ponto de vista teórico o problema de estimação em três situações:

- Quando se conhecem os parâmetros dos itens, estimam-se apenas as habilidades;
- Quando se conhecem as habilidades dos respondentes, estimam-se os parâmetros dos itens; e
- Quando não se conhecem nem os parâmetros dos itens nem as habilidades dos respondentes, estimam-se ambos simultaneamente.

Considerando-se o caso mais comum, no qual nem os parâmetros nem as habilidades são conhecidos, o problema da estimação é determinar o valor de θ_j para cada examinado e os parâmetros que compõe cada item do teste. Neste caso, o número de parâmetros a serem estimados simultaneamente pode ser extremamente grande, levando a um enorme esforço matemático. Para contornar este problema, são utilizados alguns métodos tais como Máxima Verossimilhança; Máxima Verossimilhança Marginal; Máxima Verossimilhança Conjunta; Máxima Verossimilhança Condicional e Bayesiano (EMBRETSON; REISE, 2000). Segundo Andrade, Tavares e Valle (2000) as equações de estimação, por estes métodos assumem uma complexidade matemática que não possibilita solução analítica, e por isso devem ser resolvidos por meio de procedimentos numéricos iterativos. Os métodos iterativos mais usados na TRI são o algoritmo Newton-Raphson e o método “Scoring” de Fisher, além de outros, como os métodos de quadratura e o algoritmo EM.

No início, as estimações na TRI eram feitas pelos métodos da Máxima Verossimilhança Condicional, utilizada por Rasch (1960), onde a função verossimilhança está condicionada ao número de escores corretos, e pelo método da Máxima Verossimilhança Conjunta, em que os parâmetros dos itens e a habilidade do respondente são estimadas simultaneamente (Lord 1974). No entanto, esta última, envolve um número muito grande de parâmetros a serem estimados simultaneamente e, conseqüentemente, grandes problemas computacionais. Na

tentativa de solucionar este problema, Bock e Lieberman (1970) introduziram o método da Máxima Verossimilhança Marginal em que as estimativas são realizadas em duas etapas. Na primeira são estimados os parâmetros dos itens, assumindo uma distribuição para as habilidades e, na sequência, assumindo-se os parâmetros dos itens conhecidos, estima-se as habilidades. Apesar de minimizar o problema anterior, este método ainda necessitava de um intenso esforço analítico que foi facilitado por Bock e Aitkin (1981) que propuseram uma modificação no método utilizando o algoritmo EM que é um processo iterativo para determinação de estimativas de máxima verossimilhança de parâmetros de modelos de probabilidade na presença de variáveis aleatórias não observadas.

Mislevy (1986), Swaminathan e Gifford (1982, 1985, 1986) estabelecem distribuições a priori para os parâmetros dos itens e para as habilidades, a chamada Estimativa Bayesiana, eliminando alguns problemas, como estimação de parâmetros impróprios e não convergentes, encontrados com os métodos de Máxima Verossimilhança conjunta e marginal. Segundo Francisco (2005), a estimação por máxima verossimilhança apresenta problemas na estimação de itens que são respondidos corretamente, ou incorretamente, por todos os indivíduos, e também das proficiências de indivíduos que responderam corretamente, ou incorretamente, a todos os itens. Além disso, há a possibilidade de que as estimativas dos parâmetros dos itens caiam fora do intervalo esperado, tal como valores de a_i negativos, ou valores de c_i fora do intervalo $[0,1]$. A metodologia bayesiana apresenta uma solução em que estes problemas são contornados.

Os equacionamentos matemáticos destes métodos de estimação podem ser encontrados em Andrade, Tavares e Valle (2000), Baker e Kim (2004) Van Der Linden e Hambleton (1996).

No presente trabalho deseja-se estimar tanto os parâmetros dos itens quanto as habilidades, e utilizou-se o método da máxima verossimilhança marginal onde é realizada primeiramente a estimação dos itens e, posteriormente, das habilidades.

2.4.6 Softwares

Existem vários programas que executam os procedimentos da TRI, tais como, TESTFACT (WILSON; WOOD; GIBBONS, 1991) que pode ser usado para avaliar a dimensionalidade de um conjunto de itens dicotomizados e para estimar parâmetros de itens no modelo ogiva

normal unidimensional ou multidimensional. O BILOG-G (MISLEVY; BOCK, 1990) e BILOG-MG (ZIMOWSKI ET AL., 1996), são os softwares mais populares, cuja diferença básica entre eles é que, o primeiro permite analisar as respostas de apenas uma população, enquanto que o segundo, é aplicado para mais de um grupo de respondentes, no entanto, ambos permitem a estimação de modelos logísticos de 1, 2 ou 3 parâmetros e exigem dados dicotomizados. O MULTILOG (THISSEN, 1991) e o PARSCALE (MURAKI; BOCK, 1993) permitem a estimação de conjuntos de itens que possuam dados dicotômicos ou politômicos.

O Software utilizado no presente trabalho é o BILOG-MG, que realiza a calibração dos itens, através do método da Máxima Verossimilhança Marginal (BOCK e AITKIN, 1981) e o método EAP bayesiano de estimação por maximização da distribuição marginal *a posteriori* (MISLEVY; BOCK, 1990). Tanto o BILOG-G como os demais programas citados anteriormente assumem que os respondentes representam uma amostra aleatória de uma população de proficiências que pode ser assumida como tendo ou uma distribuição normal padrão, ou ainda uma distribuição discreta arbitrariamente especificada pelo usuário, ou ainda uma distribuição empírica, a ser estimada conjuntamente com os parâmetros dos itens.

2.5 Considerações Finais do Capítulo

Neste capítulo, foram descritos os três eixos que sustentam a presente pesquisa: usabilidade, comércio eletrônico e teoria da resposta ao item. Em relação a usabilidade, verifica-se que existem inúmeras terminologias para defini-la, entretanto, a maioria converge para as mesmas características : facilidade de uso, eficiência, efetividade e satisfação. Sendo assim, adota-se nesta pesquisa a definição de usabilidade como a capacidade que um sistema interativo oferece a seu usuário, em determinado contexto de operação, para realização de tarefas de maneira eficaz, eficiente e agradável (ISO 9241-11, 1998). Esta característica se aplica a todo tipo de interação, mas recebe atenção especial na aplicação comercial, como é o caso do *e-commerce*. Uma vez que a rápida proliferação de *sites* comerciais na web, torna este mercado cada vez mais competitivo e de certa forma torna o usuário menos tolerante a *sites* menos eficientes, ou seja, que tenham baixa usabilidade.

Para a avaliação de usabilidade, há de se considerar alguns atributos ou critérios ergonômicos, a partir dos quais é possível desenvolver ou aplicar instrumentos para avaliação da usabilidade. Os critérios ergonômicos descritos e explorados neste trabalho foram

principalmente os de Bastien e Scapin (1993), devido a sua abrangência, detalhamento e acessibilidade.

As técnicas de avaliação de usabilidade podem ser classificadas basicamente, pelo envolvimento ou não do usuário. Evidentemente todas têm suas vantagens e desvantagens assim como melhor adequação a determinados contextos. O Quadro 2.2 resume algumas das técnicas utilizadas para avaliação de ergonomia de interface e usabilidade, com algumas das suas vantagens e desvantagens.

Observa-se que a maioria delas possui como desvantagem a subjetividade e a necessidade de vários recursos. Por outro lado, a Teoria da Resposta ao Item (TRI) se mostra como uma técnica capaz de conferir objetividade e simplicidade na aplicação. Baseada em modelos matemáticos capazes de avaliar requisitos e gerar uma escala comparativa, a TRI apresenta-se como uma alternativa interessante para avaliar usabilidade em *sites* comerciais. Como vantagens pode-se citar a confiabilidade e objetividade do método bem como sua robustez e dependendo da construção do conjunto de itens pode ser bastante abrangente. Como desvantagem pode-se citar sua complexidade conceitual e a necessidade de se ter uma amostra inicial relativamente grande para sua validação. Entretanto, após sua validação a aplicação pode ser feita de forma bastante rápida, sem necessitar de conhecimentos aprofundados em usabilidade, podendo assim servir de guia para avaliações mais completas com redução de custos.

3 PROCEDIMENTOS METODOLÓGICOS

Este capítulo apresenta os procedimentos metodológicos utilizados para elaboração do construto – conjunto de itens de avaliação de usabilidade, assim como sua estruturação. É descrita também a técnica utilizada para a coleta de dados. E por fim, o tratamento dos dados através da Teoria da Resposta ao Item.

3.1 Classificação da Pesquisa

O desenvolvimento do estudo é baseado no aprofundamento da estrutura conceitual sobre medidas de usabilidade com o objetivo de gerar conhecimento sobre esta área em paralelo a área de comércio eletrônico. E a partir daí desenvolve um construto de avaliação de usabilidade baseada em questionário e em coleta de dados em uma amostragem de *sites* de *e-commerce* criando assim uma escala de medida padronizada de usabilidade baseada nos modelos matemáticos da Teoria da Resposta ao Item.

Em relação à natureza, a pesquisa é do tipo teórico aplicada, pois visa propor, aplicar e validar uma ferramenta para avaliação de usabilidade em *sites* comerciais. A abordagem da pesquisa é do tipo quantitativa, pois está relacionada à quantificação de dados obtidos mediante pesquisa e baseada em técnica estatística para o tratamento dos dados. Quanto aos objetivos, uma pesquisa pode ser classificada como exploratória, descritiva ou explicativa (SANTOS, 1999; GIL, 2002). Por ter como propósito inicial colher informações para um melhor entendimento do assunto a pesquisa pode ser classificada como exploratória.

Quanto aos procedimentos técnicos, as pesquisas podem ser classificadas como bibliográfica, experimental, *ex-post facto*, estudo de corte, levantamento, estudo de campo, estudo de caso, pesquisa ação e pesquisa participante (GIL 2002). Este estudo compreende os procedimentos técnicos de pesquisa bibliográfica e experimental.

3.2 Elaboração do Conjunto de Itens

A revisão da literatura sobre os três eixos que informam a pesquisa apresentada no capítulo anterior, serviu de arcabouço teórico para a criação e estruturação do conjunto de itens.

Do ponto de vista de usabilidade, buscou-se analisar questões e/ou recomendações ligadas a páginas da web, mais especificamente páginas comerciais. Existe na literatura grande número de critérios e recomendações gerais de usabilidade baseados na interação Humano-Computador de uma forma generalista, incluindo avaliação de softwares, hardware e *sites* em geral. Recomendações específicas para *e-commerce* são raras. As recomendações usadas como base para elaboração do questionário além dos critérios ergonômicos gerais citados no capítulo 2, foram:

- *E-commerce User Experience 207: Guidelines for e-commerce sites*, Nielsen *et al* (2001) onde os autores definem 207 diretrizes para *sites* de *e-commerce*, baseados em testes com usuários realizado nos EUA e na Europa.
- *Research-Based web design and usability Guidelines* desenvolvido pelo departamentos de saúde e serviços humanos dos EUA (HHS, 2006).
- *Prioritizing web usability*, de Nielsen e Loranger (2006) que é uma atualização e ampliação das diretrizes publicadas por Nielsen *et al* (2001).

A revisão bibliográfica sobre o mercado de *e-commerce* serviu de base para gerar conhecimento sobre as modalidades existentes neste, e principalmente sobre o comportamento do consumidor, verificação de suas prioridades e fatores que podem influenciar a interação. Servindo assim de base para construção dos itens e da estrutura do conjunto.

A Teoria da Resposta ao Item (TRI), como visto no capítulo dois, é uma ferramenta robusta para análise de itens e construção de escalas. Entretanto, sua operacionalização requer a consideração de alguns pressupostos, que foram considerados nesta fase da construção dos itens, entre eles:

- unidimensionalidade: segundo Singh (2004) a dimensionalidade de um conjunto de itens está associada ao número de fatores necessários para explicar a variabilidade dos dados, ou seja, o número de traços latentes envolvidos no construto. Existem modelos na TRI capazes de avaliar um conjunto de itens com mais de uma dimensão, no entanto a complexidade envolvida limita sua operacionalidade. Desta forma, formulou-se o pressuposto inicial de que o conjunto de itens final tenha unidimensionalidade, ou seja, que esteja de fato medindo apenas um traço latente: a usabilidade em *sites* de *e-commerce*.

- interdependência local: esta característica supõe que as respostas aos itens do conjunto sejam independentes, ou seja, nenhuma resposta depende de outra resposta, dada uma certa proficiência.
- Objetividade: este pré-suposto garante que um item tenha apenas uma resposta.
- Simplicidade e clareza: garante o entendimento do item por parte do respondente. Segundo Pasquali (1998), um item deve expressar uma única idéia e deve ser inteligível para todos os estratos da população- meta.

Sendo assim, após este levantamento bibliográfico e a consideração dos pressupostos citados, partiu-se para a elaboração do conjunto de itens.

3.2.1 Conceitos utilizados

Como visto no capítulo dois, os conceitos envolvidos em uma pesquisa representam os termos abstratos desta, traduzidos em termos práticos pelos itens. Na presente pesquisa, os conceitos considerados na construção do conjunto de itens estão relacionados aos objetivos básicos de um usuário de *e-commerce* que são: visualização inicial dos elementos de uma página, navegação por esta, busca por produtos, busca por informações e efetivação da compra. Sendo assim, os conceitos abordados na presente pesquisa, para elaboração dos itens foram descritos como:

- design da página;
- navegação;
- busca e informações; e
- suporte a tarefa de compra.

3.2.1.1 Design da página

Este conceito está relacionado principalmente com o aspecto visual do site e à eficiência na transferência de informação visual. Segundo ISO 9241-3 (2000) eficiência neste contexto significa proporcionar ao usuário capacidade para detectar e reconhecer o objetivo visual do site, rapidamente e sem desconforto visual ou de ordem cognitiva.

Uma página com um bom design tem como função manter a atenção do usuário nos primeiros instantes da interação, respeitando seus objetivos, sendo simples e agradável.

Dentro deste aspecto, encontram-se várias recomendações como as referentes à primeira impressão gerada por um site. Segundo Nielsen e Loranger (2006) o tempo médio gasto em uma página inicial (*homepage*) é em média de 30 segundos, sendo assim, todas as mensagens relevantes ao consumidor devem ser expostas de forma clara, simples e direta, tais como:

- O site a que o usuário chegou;
- Os benefícios que a empresa oferece;
- Algo sobre a empresa e seus produtos mais recentes ou novos desenvolvimentos; e
- As opções para os usuários e como chegar à seção mais relevante para eles.

HHS (2006) recomendam que o conteúdo principal da página inicial deve ser visível na primeira página, isto é, sem necessidade de rolagem, e deve possuir quantidade limitada de textos densos.

O design de uma página está alinhado também à experiência dos usuários condicionando assim padrões, ou seja, o usuário espera que determinados elementos estejam posicionados em determinado lugar e que funcionem da forma com a que eles estão acostumados na maioria dos *sites*. Segundo Nielsen e Loranger (2006), os padrões de design asseguram que usuários:

- saibam quais recursos esperar;
- reconheçam a aparência desses recursos na interface;
- saibam onde encontrar esses recursos no site e na página;
- saibam como operar cada recurso para alcançar seus objetivos;
- não precisem adivinhar o significado de elementos de design desconhecidos;
- não deixem que recursos importantes passem despercebidos porque eles não prestaram atenção a um elemento de design que não é padrão; e
- não obtenham surpresas desagradáveis quando algo não funciona como o esperado.

Este conceito abrange também: legibilidade, layout, rolagem ou paginação, multimídia, anúncios publicitários.

A legibilidade está relacionada às características lexicais das informações dispostas na tela, tais como: brilho do caractere, contraste letra/fundo, tamanho da fonte, espaçamento entre palavras, espaçamento entre linhas, espaçamento de parágrafos, comprimento da linha, entre outras, que podem dificultar ou facilitar a leitura das informações (BASTIEN; SCAPIN, 1993).

O layout da página está associado à estrutura da página no que diz respeito à ordem dos itens no espaço disponível bem como sua posição, alinhamento e estrutura de cores. Um bom

layout permite aos usuários comparar dois ou mais itens sem precisar memorizar um deles enquanto navega em outras páginas ou até mesmo na mesma página, ou seja, um bom layout permite ao usuário identificar claramente similaridades, diferenças, tendências e relações (SPOOL, ET AL., 1997; WILLIAMS, 2000; HHS, 2006)

Com relação à rolagem ou paginação, segundo HHS (2006), o projetista deve decidir se deseja criar páginas longas, que requerem rolagem, ou se deseja páginas curtas que requerem frequente movimentação entre páginas (paginação). Esta decisão é função, principalmente, da quantidade de informação e/ou produtos existentes, do público-alvo e do tipo de atividade a ser executada. Entretanto, esta não é uma questão que possua um consenso na literatura. Alguns autores defendem a utilização de rolagem, principalmente em páginas onde exista leitura (BYRNE, ET AL, 1999; CAMPBELL; MAGLIO, 1999; PIOLAT; ROUSSEY; THUNIN, 1998; SPOOL, ET AL, 1997; SPYRIDAKIS, 2000). Outros defendem a utilização de paginação principalmente quando o tempo de resposta do sistema do usuário for rápido (NIELSEN, 1997; PIOLAT; ROUSSEY; THUNIN, 1998). A utilização de páginas curta com pouca rolagem é defendida por autores como Detweiler e Omanson, (1996); Lynch e Horton (2002); Nielsen (1996); Spool *et al.* (1997); Spyridakis (2000). Para estes, na maioria dos *websites*, os usuários lidam melhor com páginas de informação curtas e bem organizadas ao invés de páginas longas onde a rolagem pode tomar muito tempo do usuário.

A rolagem horizontal é considerada por alguns autores como extremamente nociva à usabilidade de uma página, pois a maioria dos usuários a considera lenta e tediosa (BERNARD; LARSEN, 2001; LYNCH; HORTON, 2002; NILSEN; TAHIR, 2002; SPYRIDAKIS, 2000; WILLIAMS, 2000, NIELSEN; LORANGER, 2006).

Em seus estudos com usuários de *e-commerce*, Nielsen e Loranger (2006), utilizaram uma amostragem de 3992 páginas que eram mais longas que a janela do navegador e concluíram que mais da metade dos usuários não rolaram a página. Em seus estudos, somente 42% dos usuários viram qualquer informação na segunda tela cheia e somente 14% visualizaram mais de duas telas cheias.

3.2.1.2 Navegação

Segundo HHS (2006, p82) a “navegação refere-se ao método usado para encontrar informações dentro de um web site.” Ligado assim, a eficiência e eficácia na execução de

tarefas no site, acrescentando as características visuais: previsibilidade (respeitando convenções da web e os modelos mentais dos usuários), confiabilidade e *feedback*. Para Cybis (2007) uma interface ergonômica deve antecipar-se e reagir às ações dos usuários, convidando-o, apoiando-o durante suas ações e fornecendo-lhe respostas adequadas quanto ao andamento e ao resultado dessas ações. Sendo assim, a navegação é um meio por onde o usuário necessariamente deve passar para chegar onde deseja, e quanto mais eficiente e previsível ela for maior a usabilidade de um site.

Quando a navegação muda drasticamente de uma página para outra, os usuários precisam desviar a atenção do uso do site para descobrir como utilizá-la. *Sites* sem uma boa e segura navegação tornam os usuários inseguros e hesitantes (NIELSEN; LORANGER, 2006).

Recomendações sobre feedback podem ser encontradas em Apple (1992) onde é argumentado que a manipulação direta dá às pessoas a sensação de que estão controlando os objetos representados no computador; o objeto manipulado deve permanecer visível enquanto estiver sendo realizada uma operação sobre ele e o efeito dessa operação deve ser imediatamente notado.

Confiança e credibilidade são questões importantes na web, em que mesmo a maior empresa só existe aparentemente virtualmente. Segundo Nielsen e Loranger (2006) a empresa mais fraudulenta e aética pode parecer tão convincente quanto uma com um longo histórico em serviços comunitários e relacionamento honesto com os clientes. Desta forma, para um website ser bem-sucedido, o usuário precisa saber quem está por trás dele, se é confiável ou não. Por isso segundo os autores a maioria dos usuários busca informações sobre a empresa e quando não a encontram ficam frustradas e se questionam se a empresa é de confiança se possui endereço e telefone reais ou se é apenas uma empresa de fachada.

Alguns elementos de design são ativados durante a navegação do usuário sem que ele solicite, como é o caso dos anúncios publicitários *pop-ups* ou a abertura de novas janelas, que podem confundir o usuário além de serem desagradáveis, como argumentam HHS (2006), Nielsen e Norman (2001) e Nielsen e Loranger (2006).

3.2.1.3 Busca e informações

Este aspecto abrange a funcionalidade da ferramenta de busca e a apresentação do resultado da busca que inclui a forma de apresentar as informações referentes aos produtos.

A busca dentro de um site é uma etapa fundamental da interação do usuário com a empresa de *e-commerce*, o usuário gera grande expectativa sobre sua funcionalidade. Usuários querem estar aptos a usar o resultado da pesquisa para continuar resolvendo seus problemas. Quando os usuários ficam confusos com o resultado da busca, ou não encontram imediatamente o que eles estão procurando, frustram-se (AMENTO *ET AL*, 1999; BAILEY, KOVANI, 2004; DUMAIS, CUTRELL, CHE, 2001; NIELSEN, 2001; NIELSEN *ET AL*, 2001; POLLOCK, HOCKLEY, 1996; ROSENFELD, MORVILLE, 2002; SOOP *ET AL*, 1997)

Segundo Nielsen e Loranger (2006), as três coisas que os usuários esperam de um sistema de busca são:

- uma caixa em que possam digitar palavras;
- um botão rotulado Pesquisa em que clicam para executar a busca; e
- uma lista dos principais resultados apresentada de forma linear.

Ainda segundo os autores, se o usuário não puder localizar o produto, ele não vai comprá-lo. A capacidade de se movimentar em um Website é extremamente importante para a usabilidade, mas os principais componentes Buscar e Localizar são responsáveis por mais de um terço das dificuldades dos usuários que fazem isso.

O resultado da busca e as informações contidas nele são importantíssimas no processo de compra em *sites* comerciais. Uma compra online pode provocar no usuário certo grau de incerteza e de insegurança quando comparada a mesma compra em meios tradicionais offline – como lojas por exemplo. Embora o site possua um design e uma navegação consistente, se o usuário não conseguir encontrar as informações que ele deseja sobre o produto que ele está procurando dificilmente a compra será efetivada. As informações sobre produtos podem ser as mais diversas possíveis, desde descrições técnicas necessárias a determinados produtos até imagens em três dimensões, entendendo-se aqui como informações sobre produtos toda informação que possa influenciar na decisão de compra.

Em lojas online, não há um produto tangível para segurar, nem vitrines de produtos para inspecionar. Na maioria dos casos, não há vendedor ao qual se possa pedir ajuda. Desta forma, os *websites* devem preencher essa lacuna com boas e confiáveis informações. Numa época em que multiplicam-se as fraudes na Internet, os clientes precisam de garantias extras de que as empresas são confiáveis antes de se comprometer com uma venda. Portanto, é essencial ser simples e claro, principalmente sobre informações de produtos (NIELSEN; LORANGER, 2006).

Deve-se observar toda a funcionalidade de uma ferramenta de busca, assim como a contextualização dos elementos apresentados, pois a ação da venda, somente será concretizada caso o consumidor seja atraído e sinta-se satisfeito na utilização dos componentes do site e na navegação por suas páginas, tendo facilidade em encontrar o que procura e não opondo-se a disponibilizar informações solicitadas para a efetivação de sua compra (SHNEIDERMAN, 1998).

3.2.1.4 Suporte à tarefa de compra

Este conceito envolve aspectos visuais e experiência do usuário, no entanto com objetivo final de efetivação da compra, como preenchimento de formulários, facilidade de compra, previsão de erros, segurança, entre outros.

Segundo Bastien e Scapin (1993), uma interface gráfica deve ser convidativa e auxiliar o usuário em sua tarefa, no caso de formulários, ela deve conter informações sobre o preenchimento deste, sobre as entradas esperadas, como seu nome, sua descrição, seu formato e as unidades de medida.

Um formulário mal projetado pode ocasionar erros ou confusão visual e ser responsável por vendas perdidas. Entretanto, por mais bem projetado que possa ser um formulário, podem surgir erros decorrentes de entrada de dados equivocadas por parte do usuário, para isso é importante ter um bom processo de gestão de erros. Para Bastien e Scapin (1993), a gestão de erros está ligada a todos os mecanismos que permitem evitar ou reduzir a ocorrência de erros e que favoreçam sua correção (proteção contra erros, qualidade das mensagens de erros e correção dos erros).

3.2.2 Organização do conjunto de itens

Após a etapa de formulação dos itens, baseados nas definições dos conceitos juntamente com revisão bibliográfica e as recomendações encontradas, obteve-se um conjunto de 73 itens (Apêndice B) com resposta dicotomizadas, ou seja, apresenta ou não a característica (0 ou 1).

A sequência deste conjunto de itens tentou seguir a ordem de passos seguida por um usuário em um site de *e-commerce*, ou seja:

- entrar pela página inicial;

- buscar determinado produto;
- retornar à página inicial;
- buscar outro produto;
- efetivar a compra do produto; e
- retornar à página inicial.

Segundo Nielsen e Loranger (2006), 60% dos usuários chega a uma página por intermédio de um site de busca, portanto acessam páginas internas. Desta forma, as etapas de retorno à página inicial têm por objetivo simular esta situação de forma a testar a orientação que o site passa ao usuário quanto ao restante de seu conteúdo.

Inicia-se o procedimento pela página inicial pelo fato de se estar simulando a avaliação de determinados *sites* específicos escolhidos aleatoriamente, seguindo assim uma avaliação linear para todos os *sites* testados. Na primeira etapa, é possível avaliar principalmente aspectos relacionados ao design da página. Na passagem de uma etapa para outra é possível testar a navegação do site. A tarefa de busca avalia a ferramenta de busca do site e as informações contidas, e por fim a simulação da efetivação da compra do produto avalia esta parte final da interação site-usuário. Sendo assim, a sequência dos itens tentou seguir esta ordem de tarefas de forma a tornar a avaliação mais eficiente.

3.2.3 Teste piloto

O primeiro conjunto de itens (Apêndice B) foi testado em uma amostra de cerca de 20 *sites* de *e-commerce*. O objetivo foi verificar o comportamento dos itens e eventualmente identificar problemas relacionados à sua operacionalidade, possível dependência entre itens ou algum desrespeito aos pré-supostos da Teoria da Resposta ao Item.

Neste teste foram eliminados 29 itens, restando, portanto 44 itens a serem testados na amostragem de 361 *sites*. Dos 29 itens eliminados, as principais causas da eliminação foram a subjetividade da avaliação do mesmo e a dependência entre itens. A primeira, além de inserir imprecisão nos resultados, pelo fato da análise ter sido feita por apenas um usuário, ainda aumentaria significativamente o tempo de análise. A segunda causa contraria o pressuposto básico da TRI de interdependência local. O teste piloto serviu também para colocar os itens restantes (44) em uma ordenação que reduzisse o tempo de avaliação de cada site, tornando-a a mais linear possível, sem idas e vindas desnecessárias no site durante a avaliação.

3.2.4 Itens resultantes

Os 44 itens resultantes da elaboração inicial e do teste piloto, juntamente com suas descrições podem ser visualizados no Quadro 3.1. Observa-se que a sequência dos itens prioriza a linearidade na aplicação da avaliação, ou seja, os doze primeiros itens podem ser avaliados logo no primeiro acesso ao site. Os demais, até o item 31, são avaliados durante a navegação à procura de determinado produto. As características abordadas entre os itens 32 e 38 são avaliadas durante o preenchimento do formulário de compra, e os últimos itens são verificados durante a navegação de volta a página inicial, o que resulta em menos tempo de avaliação dos *sites*.

Item	Descrição
01	Ao abrir a página inicial, não há abertura de janelas pop-up?
02	A página inicial 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?
03	A página inicial possui menos de 2,3 páginas de rolagem?
04	As palavras clicáveis (coloridas ou sublinhadas) apresentam uma forma diferenciada quando são selecionadas?
05	O layout da página é adaptável ao tamanho da janela, pelo menos horizontalmente (ou seja, não há rolagem horizontal)?
06	As opções principais do site estão visíveis? Ou seja, não há necessidade de passar o mouse sobre uma área gráfica para ver opções ocultas?
07	O site possui a opção “ajuda”?
08	Na área principal do site, existem textos com no máximo quatro cores diferentes?
09	Na área principal do site, existem textos com no máximo três fontes diferentes?
10	A disposição dos objetos de interação de uma caixa de diálogo segue uma ordem lógica?
11	Os rótulos de campos começam com uma letra maiúscula e as letras restantes são minúsculas?
12	O site possui opção de acesso em outras línguas?
13	É possível encontrar informações sobre a empresa (endereço físico, telefone etc)?
14	Os títulos dos textos estão alinhados à esquerda?
15	Os parágrafos de texto são separados?
16	As palavras aparentemente clicáveis são de fato clicáveis ?
17	Os títulos de telas, janelas e caixas de diálogo estão no alto, centrados ou justificados à esquerda?
18	Todas as páginas possuem um campo de busca?
19	Os resultados de busca permitem classificação por outros critérios, além de custo?

(Continua...)

(Continuação...)

Item	Descrição
20	Listas longas apresentam indicadores de continuação, de quantidade de itens e de páginas?
21	O preço de um produto consta ao lado da imagem ou do link do produto?
22	Existe uma orientação ao usuário quanto ao restante do site?
23	A maioria dos produtos possui informações sobre eles?
24	O tamanho da fonte dos textos é de dez pontos ou mais?
25	É possível ampliar as fotos dos produtos para visualizar detalhes?
26	Em produtos em que existam mais de uma perspectiva, é possível visualizar todas as perspectivas?
27	Os grupos de botões de comando estão dispostos em coluna e à direita, ou em linha e abaixo dos objetos aos quais estão associados?
28	Quando há rolagem, não existem elementos de design (na tela inicial) que pareçam com marcadores de final de página?
29	Todos os campos e mostradores de dados possuem rótulos identificativos?
30	O botão de finalização de compra está no final da lista?
31	É possível saber os custos totais antes de fazer cadastro?
32	Os campos destinados a formulários de preenchimento pelo usuário estão agrupados linearmente, evitando espaços desnecessários?
33	No preenchimento de um formulário, as opções que não são válidas ou não estão disponíveis estão visivelmente desativadas, prevenindo erros?
34	No preenchimento de um formulário, é informada a forma de preenchimento?
35	Os dados obrigatórios são diferenciados dos dados opcionais de forma visualmente clara?
36	É possível efetivar uma compra sem realizar um cadastro (que inclua login e senha)?
37	O sistema emite sinais sonoros quando ocorrem problemas na entrada de dados?
38	As mensagens de erro estão isentas de abreviaturas e/ ou códigos gerados pelo sistema operacional?
39	Todas as páginas possuem os mesmos <i>layouts</i> e exibem ao usuário as mesmas características?
40	O logotipo da empresa está no canto superior esquerdo em todas as páginas do site?
41	Existe um link de um único clique que conduz à página inicial?
42	Qualquer ação do usuário pode ser revertida através da opção DESFAZER?
43	O site permite navegação em suas páginas em apenas uma janela, ou seja, não há abertura de novas janelas em meio à navegação?
44	Os links já visitados mudam de cor?

Quadro 3.1 - Os itens e suas descrições.

A classificação dos 44 itens segundo conceitos de usabilidade abordados pode ser visualizada no Quadro 3.2.

Conceito de usabilidade	Itens
Design da página	02, 03, 05, 06, 08, 09, 10, 11, 14, 15, 17, 24, 27, 28.
Navegação	01, 04, 07, 12, 13, 16, 20, 22, 37, 38, 39, 40, 41, 42, 43, 44.
Busca e informações	18, 19, 21, 23, 25, 26, 29, 31.
Suporte à tarefa de compra	30, 32, 33, 34, 35, 36.

Quadro 3.2 - Classificação dos itens conforme o conceito de usabilidade abordado.

Observa-se que a maioria dos itens refere-se aos conceitos relativos à navegação, com dezesseis itens, e ao design da página, com quatorze itens. Juntos, estes conceitos representam 68% do conjunto de itens.

3.3 População e Amostra

A amostra utilizada na presente análise foi formada por uma amostragem da população de *sites* de *e-commerce* de origem brasileira.

Em termos estatísticos, segundo Freund e Simon (2000), a população consiste de todas as observações concebíveis (ou hipoteticamente possíveis) de um determinado fenômeno. Os elementos que compõem uma população podem ser indivíduos, empresas, produtos manufaturados, preços ou qualquer coisa que possa ser mensurada, contada ou ordenada. Desta forma, a população observada nesta pesquisa comporta todos os *sites* de *e-commerce* de origem brasileira, ou seja, qualquer site que comercialize algum tipo de produto. Amostra é por sua vez uma parte da população, selecionada através de uma amostragem que teoricamente compreende técnicas convenientemente definidas para selecionar uma amostra representativa de uma população para fazer generalizações sobre a população. Amostragem aleatória simples é um tipo de amostragem onde são escolhidos elementos amostrais de forma que cada elemento amostral tem a mesma chance de ser escolhido. Para obter uma amostra é preciso estabelecer uma população alvo (STEVENSON, 1981; MILONE, 2004).

Foram analisados 361 *sites* de *e-commerce* (apêndice A), abrangendo vários gêneros, escolhidos de forma aleatória utilizando *sites* de busca e pesquisas na internet. Durante o processo de amostragem tomou-se o cuidado de garantir que a mesma fosse a mais diversificada possível e ao mesmo tempo garantir sua representatividade. Para isso, além de *sites* com os mais diversos gêneros de produtos comercializados, observou-se a variedade de

estilos de design, contemplando desde algo relativamente primitivo até excessivamente elaborado, o que não necessariamente implica em melhor usabilidade, apenas garante a diversidade, pré-requisito da Teoria da Resposta ao Item. O Quadro 3.3 demonstra a composição da amostra segundo o tipo de produto comercializado pelo site:

Produto	Sites (quantidade)	Sites (%)
Acessórios automotivos	9	2,49
Alimentos e Bebidas	20	5,54
Artesanato	7	1,94
Artigos esportivos	21	5,82
Bolsas e acessórios	32	8,86
Brinquedos	16	4,43
Calçados	8	2,22
Cosméticos	7	1,94
DVD	8	2,22
Eletrodomésticos	14	3,88
Eletrônicos	27	7,48
Floricultura	6	1,66
Informática	20	5,54
Instrumentos musicais	7	1,94
Livraria	20	5,54
Loja de departamentos	19	5,26
Medicamento	5	1,39
Móveis	10	2,77
Outros	19	5,26
Papelaria	7	1,94
Perfumes	13	3,60
Pet Shop	7	1,94
Produtos específicos/personalizados	16	4,43
Produtos exóticos/religiosos	13	3,60
Sex Shop	11	3,05
Vestuário	19	5,26
Total	361	100,00

Quadro 3.3 - Composição da amostra segundo o tipo de produto comercializado pelo site.

3.4 Coleta de Dados

O ambiente para coleta de dados foi constituído por um computador com tela de 15 polegadas. O navegador utilizado foi o Internet Explorer versão 7, por ser, segundo Nielsen e Loranger (2006), o navegador mais utilizado no mundo atualmente. A resolução da tela utilizada foi de 1024 por 768 pixels que, segundo Nielsen e Loranger (2006 p.225), “é o tamanho mais popular e a tendência aponta para 1024 por 768 como o menor tamanho de tela no futuro”.

A análise dos *sites* foi feita por apenas um usuário – o autor - tendo em vista apenas avaliação de *sites* comerciais. E consistiu-se de uma sequência de passos padronizada, com a finalidade de responder o conjunto de itens.

O primeiro passo da avaliação de um site consistiu em abrir a página inicial e identificar seus elementos visuais e sua estrutura baseado nos itens elaborados para este fim. Depois, baseado no tipo de produto que o site comercializa faz-se uma busca simples, com a finalidade de testar o resultado de seu sistema de busca e as informações sobre o produto, quando houver, e conseqüentemente responder aos itens referentes a este aspecto. Com o objetivo de testar alguns elementos navegacionais, após a busca o produto é inserido no “carrinho de compra” e em seguida retorna-se a página inicial, e a partir daí retorna-se ao carrinho de compra e simula-se a efetivação da compra respondendo assim os itens referentes a este aspecto. Após esta simulação retorna-se a página inicial através da função desfazer com o objetivo de responder os itens referentes a este aspecto.

Segundo o modelo de Schuman e Kalton (1985), a maneira de administrar ou aplicar o instrumento influencia a codificação das respostas, seu processamento e sua análise.

A codificação das respostas foi feita de forma dicotômica, representando-se 1(um) para “apresenta a característica” e 0 (zero) “não apresenta a característica”, o processamento dos dados foi feito em planilha eletrônica e a análise dos dados foi feita no software BILOG MG[®] utilizando o modelo logístico de dois parâmetros da Teoria da Resposta ao Item. Alguns *sites*, devido a suas características e/ou produto comercializado, não apresentam determinadas característica e isto não significa necessariamente redução na usabilidade. Desta forma, na coleta de dados, o site que apresentou este comportamento em determinado item, permaneceu no banco de dados e foi considerado como item não aplicável, codificado, para efeitos computacionais, com o número 9 (nove). A Teoria da Resposta ao Item permite este tratamento, devido à propriedade de interdependência dos itens e conseqüentemente, por não depender do escore final. A quantidade de respostas não aplicáveis foi de 0,42% do total da amostragem final. No contexto analisado, pelo fato destas características não serem previsíveis, tampouco restritas a um grupo e apresentaram-se raramente, decidiu-se manter os itens, que eventualmente não pudessem ser aplicáveis a todos os *sites* para preservar sua informação. Do ponto de vista da amostragem, manteve-se todos os *sites* para não comprometer a condição de aleatoriedade desta.

3.5 Tratamento dos Dados

Após o levantamento dos dados, fez-se o tratamento destes utilizando um software específico para auxiliar no procedimento de estimação dos parâmetros e do traço latente, chamado BILOG-MG[®] produzido pela *Scientific Software, Inc*, Mislevy, R. J. e Bock, R. D, (1990).

O programa exige uma sintaxe apropriada para o problema, onde é definido o tipo de análise e o modelo a ser utilizado e as especificações de entrada e saída dos dados. O modelo utilizado foi o Modelo Logístico de 2 parâmetros (ML2) (Equação 2.2). Sendo assim, no presente trabalho o parâmetro c , que representa acerto casual, não foi considerado. Este parâmetro pode ser interpretado no contexto de usabilidade em *sites* de *e-commerce* como a probabilidade de um site com baixo grau de usabilidade apresentar uma característica (descrita por um item) associada a *sites* com alto grau de usabilidade, ou seja, há probabilidade de um site apresentar boa usabilidade ao acaso? Por se entender que usabilidade é um atributo de qualidade representada pelo todo e não de forma casual, esta probabilidade não foi considerada.

Os elementos da Equação 2.2, agora no contexto de usabilidade, ficaram da seguinte forma:

U_{ij} é uma variável dicotômica que assume valor 1 (um) quando o site analisado j “responder” positivamente ao item i , ou 0 (zero) quando o site “responder” negativamente ao item i ;

θ_j representa a usabilidade (traço latente) do j -ésimo site.

$P(U_{ij} = 1|\theta_j)$ é a probabilidade de um site j com usabilidade θ_j apresentar a característica questionada no item i .

b_i é o parâmetro de dificuldade (ou de posição) do item i , medido na mesma escala da usabilidade.

a_i é o parâmetro de discriminação (ou de inclinação) do item i , com valor proporcional à inclinação da Curva Característica do Item — CCI no ponto b_i .

O programa executa a análise em três etapas, chamadas de fases 1, 2, e 3, que se caracterizam pelo tipo de tarefas realizadas em cada uma delas.

Na **fase 1** são calculadas algumas estatísticas descritivas básicas referentes a cada item, tais como: número de *sites* testados, quantidades de *sites* que responderam positivamente ao item e algumas correlações de interesses, como as correlações bisserial e ponto bisserial. Segundo Andrade, Tavares e Valle (2000) a importância dessa etapa está tanto na verificação por parte

do analista, de que a leitura dos dados foi feita corretamente, quanto na utilização destas estatísticas como valores iniciais para os processos de estimação realizados nas fases seguintes. Além disso, estatísticas como a correlação bisserial fornecem um diagnóstico preliminar dos itens, servindo, por exemplo, na identificação de itens com problemas no gabarito.

A **fase 2** é a etapa onde é produzida a calibração dos itens ou seja, nesta fase, são estimados os parâmetros dos itens (a ; b), com seus respectivos erros padrão. Segundo Embretson e Reise (2000) os resultados desta fase e da seguinte são influenciados diretamente pela escolha do modelo e consequentemente o número de parâmetros, juntamente com número de itens e o tamanho da amostragem.

A **fase 3** é a fase da estimação do grau de usabilidade (θ) dos *sites* analisados. Aqui são estimados os graus de usabilidade de cada um dos *sites* analisados, a partir dos resultados obtidos na fase anterior. Este grau de usabilidade é estimado na escala dos parâmetros dos itens. O BILOG MG[®] fornece ainda gráficos contendo algumas informações de interesse, tais como as curvas características e as curvas de informação de cada item, assim como a curva de informação total.

4 ANÁLISE DOS RESULTADOS

Nesta seção serão expostos, analisados e discutidos os resultados gerados pela TRI referente aos 44 itens elaborados e a amostragem de 361 *sites* de *e-commerce*. A sistemática de apresentação segue os passos utilizados pelo software, ou seja, verificação de estatísticas descritivas na primeira fase, calibração e estimação dos parâmetros na segunda fase e estimação dos traços latentes (usabilidade) na fase três. Os resultados destas fases são expostos e analisados. Na sequência é feita uma avaliação geral dos itens baseados nestes resultados e nos conceitos de usabilidade e por fim são identificados os itens e níveis ancora para a construção da escala proposta.

4.1 Estatística Descritiva

Na primeira fase, o software faz o reconhecimento dos dados e identifica o que é correto e o que não é correto, com isso gera algumas estatísticas descritivas tais como: número de *sites* submetidos a cada item, o número de acertos em cada item e algumas correlações de interesse, como correlação de Pearson e correlação bisserial, bastante usadas na Teoria Clássica dos Testes. Segundo Andrade, Tavares e Valle (2000) a importância desta fase está no fato de ela verificar se a leitura dos dados está sendo feita de maneira correta e estas estatísticas geradas são utilizadas posteriormente como valores iniciais para os processos de estimação realizados nas fases seguintes

A análise da correlação bisserial, verifica a existência de correlação significativa de um determinado item, com o escore bruto produzido pelo conjunto de itens. Esta verificação auxilia na identificação de itens que de fato apresentem consistência interna e se associem bem ao conjunto, ou seja, itens que apresentem correlação bisserial baixa podem ser inconsistentes com o traço latente que se deseja medir. Normalmente aceita-se valores para a correlação bisserial superior a 0,300 e com o mesmo sinal, indicando nesta caso existência de unidimensionalidade (PASQUALI, 2003; SOARES, 2005). Os resultados encontrados para a correlação bisserial, mostrados no Quadro 4.1, indicam que a maioria dos itens apresenta boa correlação bisserial, entretanto, alguns apresentam valores muito baixos, como é o caso dos itens 3 e 36.

Item	Nº Respostas	Nº Respostas corretas	Correlações	
			Pearson	Bisserial
01	361	319	0,163	0,267
02	361	267	0,247	0,334
03	361	129	0,003	0,004
04	361	272	0,335	0,458
05	361	176	0,226	0,284
06	361	324	0,194	0,329
07	361	92	0,217	0,295
08	361	204	0,207	0,260
09	361	212	0,247	0,312
10	361	235	0,312	0,402
11	361	271	0,310	0,422
12	361	12	0,111	0,266
13	361	164	0,188	0,236
14	361	319	0,245	0,401
15	351	322	0,179	0,323
16	361	167	0,384	0,481
17	361	340	0,290	0,582
18	361	312	0,180	0,283
19	350	138	0,343	0,436
20	325	181	0,292	0,367
21	360	332	0,233	0,429
22	361	167	0,341	0,428
23	361	311	0,261	0,408
24	361	114	0,095	0,124
25	358	245	0,310	0,405
26	354	65	0,263	0,383
27	361	329	0,207	0,366
28	361	338	0,154	0,300
29	361	314	0,214	0,340
30	359	344	0,116	0,258
31	359	297	0,214	0,316
32	360	298	0,100	0,148
33	357	139	0,186	0,236
34	360	221	0,333	0,424
35	360	246	0,316	0,413
36	361	42	0,004	0,007
37	361	277	0,098	0,136
38	361	337	0,246	0,474
39	361	302	0,353	0,530
40	361	317	0,363	0,587
41	361	322	0,249	0,417
42	361	282	0,361	0,506
43	361	297	0,281	0,413
44	361	20	0,121	0,246

Quadro 4.1 - Estatísticas descritivas dos itens: Correlação de Pearson e Bisserial.

Esses resultados permitem concluir que os itens 1, 3, 5, 7, 8, 12, 13, 18, 24, 30, 32, 33, 36, 37 e 44 não apresentam uma boa correlação com o escore bruto produzido pelo conjunto de

itens (correlação bisserial menor que 0,300), portanto, segundo a Teoria Clássica dos Testes, não apresenta boa discriminação para o métrica que se deseja produzir. Estes itens podem ser revistos ou eliminados, ou até mesmo mantidos para confirmação nas fases posteriores. Decidiu-se manter todos os itens para a fase posterior.

4.2 Calibração dos Itens

Na fase de calibração dos parâmetros, tem-se a estimação dos parâmetros de discriminação “a” e do parâmetro de dificuldade “b” de cada item e seus respectivos erros padrão, como pode ser verificado no Quadro 4.2. Nesta fase, o software também demonstra a convergência do processo iterativo utilizado (método EM e Newton-Raphson) e a quantidade de ciclos necessários para atingir a convergência. A possível falta de convergência indica erros com os dados ou insuficiências na amostragem.

Item	Parâmetro a	Erro Padrão	Parâmetro b	Erro Padrão	Item	Parâmetro a	Erro Padrão	Parâmetro b	Erro Padrão
01	0,597	0,155	-3,614	0,897	23	1,022	0,226	-2,109	0,367
02	0,756	0,153	-1,546	0,311	24	0,402	0,100	1,991	0,551
03	0,246	0,067	2,420	0,798	25	0,931	0,165	-0,979	0,186
04	0,979	0,175	-1,354	0,229	26	0,972	0,179	1,813	0,294
05	0,505	0,115	0,100	0,216	27	0,909	0,220	-2,904	0,584
06	0,812	0,193	-2,968	0,613	28	0,938	0,284	-3,241	0,777
07	0,648	0,127	1,798	0,369	29	0,906	0,217	-2,396	0,475
08	0,497	0,110	-0,562	0,246	30	0,752	0,247	-4,490	1,329
09	0,527	0,115	-0,716	0,255	31	0,792	0,171	-2,208	0,436
10	0,854	0,162	-0,847	0,189	32	0,490	0,125	-3,356	0,851
11	0,869	0,159	-1,461	0,268	33	0,595	0,126	0,819	0,250
12	0,790	0,212	4,651	1,134	34	1,043	0,182	-0,551	0,130
13	0,549	0,119	0,352	0,216	35	0,860	0,179	-1,036	0,215
14	0,972	0,225	-2,417	0,446	36	0,349	0,098	5,952	1,726
15	0,891	0,228	-3,032	0,641	37	0,382	0,096	-3,224	0,858
16	1,075	0,178	0,157	0,114	38	1,073	0,243	-2,881	0,510
17	1,376	0,346	-2,548	0,401	39	1,337	0,282	-1,580	0,239
18	0,755	0,183	-2,704	0,592	40	1,530	0,324	-1,742	0,244
19	1,204	0,228	0,464	0,123	41	1,178	0,279	-2,186	0,373
20	0,791	0,159	-0,278	0,158	42	1,350	0,225	-1,239	0,167
21	1,120	0,283	-2,628	0,480	43	1,056	0,200	-1,743	0,279
22	1,000	0,179	0,167	0,123	44	0,716	0,178	4,285	0,992

Quadro 4.2 - Estimação dos parâmetros dos itens.

Quanto ao grau de dificuldade ‘b’, os três itens com maior dificuldade foram, em ordem crescente, os itens 44, 12 e 36. Já os itens com menor grau de dificuldade, também em ordem crescente, foram 30, 01 e 32. É importante ressaltar que os valores referentes ao grau de dificuldade são referenciados a uma escala baseada na distribuição normal de métrica (0,1), ou seja, média 0 (zero) e desvio-padrão 1 (um). Sendo assim, os valores, como já foi descrito no capítulo 2, 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, que será descrita na próxima seção. Portanto, o item 36, que apresentou um grau de dificuldade estimado em 5,952 possui uma exigência em termos de usabilidade maior que o item 30, que apresentou um grau de dificuldade de -4,490, verifica-se assim uma comparação linear.

O parâmetro grau de dificuldade 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, determina a “qualidade” do item, ou seja, quanto maior o valor de “a” maior o grau de discriminação na região de maior informação (“b”), consequentemente melhor o item. Itens com grau de discriminação inferiores a 0,700 foram considerados pouco informativos, foram eles: 01, 03, 05, 07, 08, 09, 13, 24, 32, 33, 36 e 37. Esta afirmação pode ser confirmada através da curva de informação destes itens (Apêndice D), onde verifica-se baixíssima informação em toda a escala, inclusive na região referente ao parâmetro de dificuldade “b”. Nota-se também que todos estes itens haviam apresentado, na fase anterior, correlação bisserial baixa, o que reforça a suspeita de que estes itens não estejam realmente medindo adequadamente o grau de usabilidade dentro do modelo utilizado.

Os itens com maior grau de discriminação foram, em ordem crescente, 42, 17 e 40. O conjunto de dados apresentou uma média de 0,850 para o grau de discriminação, com um desvio padrão de 0,292, e média de -0,899 para o grau de dificuldade com um desvio padrão de 2,283.

As fórmulas para determinação do erro padrão provêm da derivada segunda da equação da TRI definida para a análise, no caso do presente trabalho a equação (2.6) – modelo logístico de dois parâmetros - em relação ao parâmetro (a, b). Uma vez que a derivada de segunda ordem reflete o quão rapidamente a inclinação da curva muda ao redor de um determinado valor, a taxa de mudança na estimativa determina o erro padrão. Na prática, a matriz informação também é calculada a partir da derivada segunda e a inversa desta matriz é interpretada como matriz de covariância dos parâmetros e a raiz quadrada dos elementos

diagonais desta matriz representam o erro padrão dos parâmetros (EMBRETON; REISE, 2000). Para mais detalhes sobre a determinação dos erros padrão na estimação dos parâmetros ver Hambleton; Swaminathan (1985); Andrade; Tavares; Valle (2000).

A análise dos resultados dos erros padrão foi realizada considerando sua magnitude e homogeneidade no conjunto. A homogeneidade é verificada quando a magnitude dos erros não sofre grande discrepância entre os itens do conjunto, podendo evidentemente varia de acordo com o valor do parâmetro. Verifica-se que no conjunto analisado os erros, tanto do parâmetro *a* quanto do parâmetro *b*, apresentam-se homogêneos. Quanto à magnitude, considerando que a amostragem segue uma distribuição normal, pode-se dizer que o valor verdadeiro do parâmetro estará entre dois erros padrão da estimativa (MONTGOMERY; RUNGER, 2003). Desta forma, por exemplo, o item 01, que apresentou um parâmetro de dificuldade (*b*) igual a -3,614 e um erro padrão igual a 0,897 poderá ter a seguinte variação: $-3,614 \pm 2 \cdot 0,897$. Ou seja, o grau de dificuldade deste item pode estar entre -5,408 e -1,820, o que representa limites aceitáveis no contexto, uma vez que existe homogeneidade no conjunto permitindo assim que a comparabilidade se mantenha.

Nesta fase são gerados também alguns gráficos que ilustram a estimação dos parâmetros e a relação entre eles, como por exemplo, o gráfico referente à curva característica do Item 40, que pode ser visualizado na Figura 4.1, onde verifica-se o parâmetro *b* = -1,742 e o parâmetro *a* = 1,530, o primeiro indica na escala de *ability* ou grau de usabilidade, onde o item possui mais informação, e pode ser interpretado como o ponto na escala onde a probabilidade de resposta positiva, por parte do site avaliado, é de 0,5. Quanto maior for o valor de *b*, maior será o grau de usabilidade exigido para que o site tenha a chance de 50% de responder positivamente o item. Esta probabilidade aumenta com o aumento do grau de usabilidade do site analisado, de acordo com a curva apresentada. Verifica-se que *sites* com grau de usabilidade igual a 0 (zero) na escala gerada, tem probabilidade de 0,93 de apresentar a característica abordada pelo item 40.

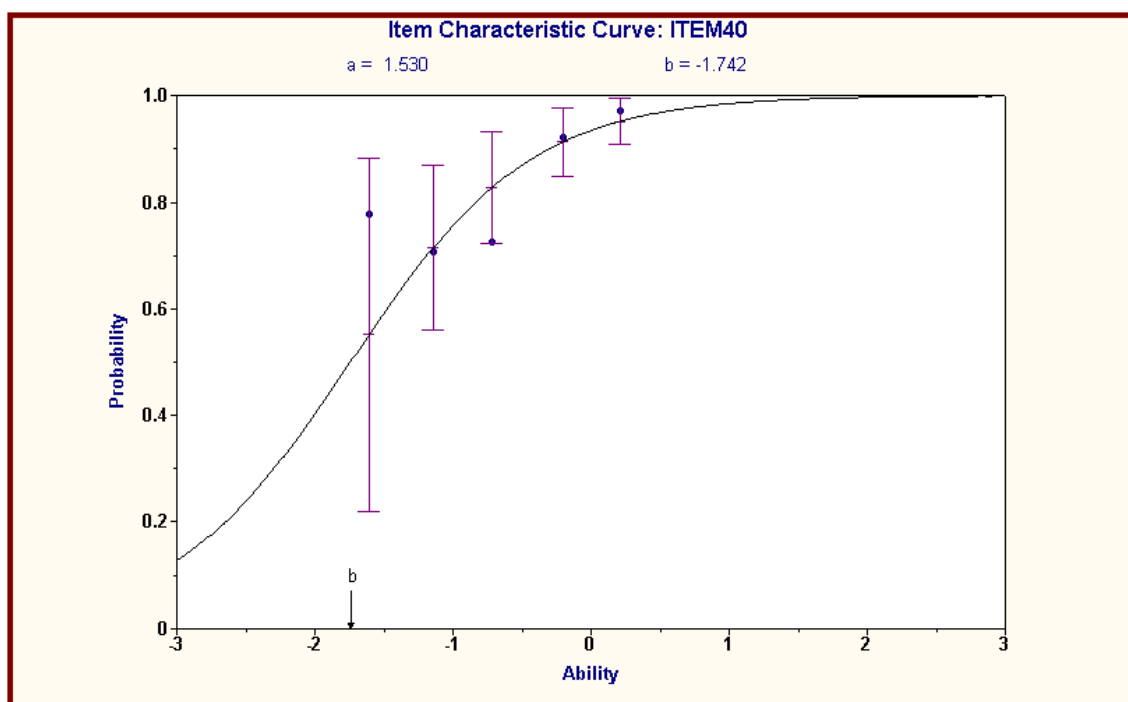


Figura 4.1 - Curva Característica do item 40

Portanto este item pode ser considerado fácil quando comparado com o item 26, por exemplo, que possui um grau de dificuldade igual a 1,813 na mesma escala, como pode ser visualizado na Figura 4.2.

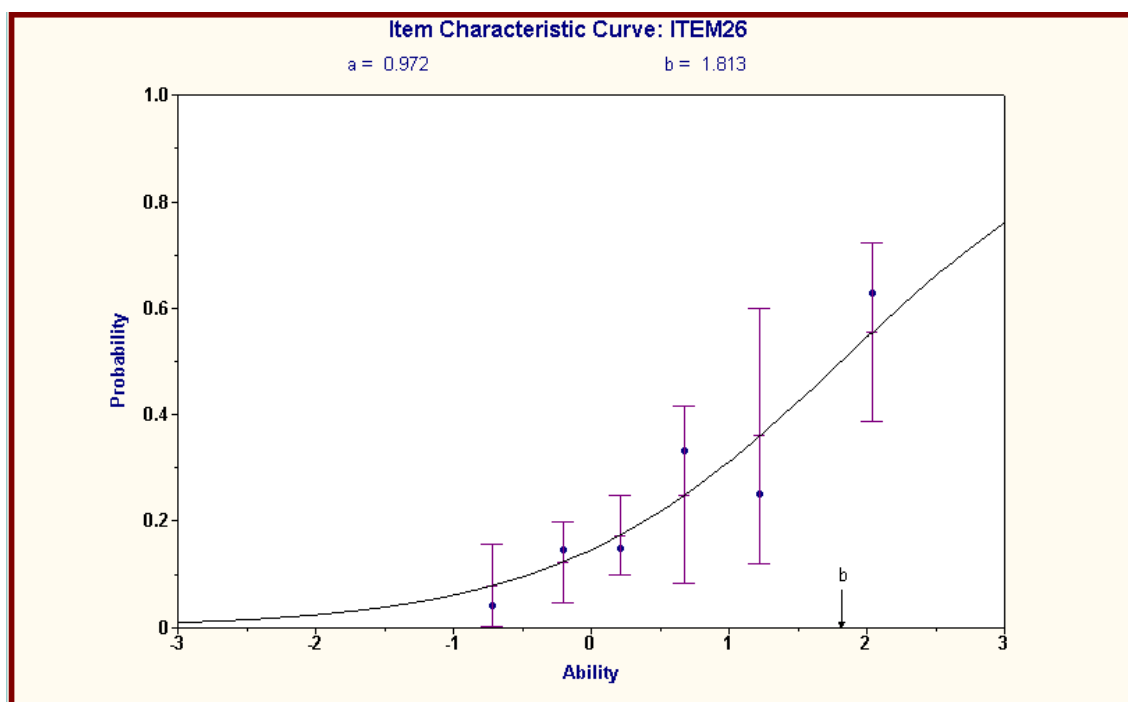


Figura 4.2 - Curva Característica do item 26.

O parâmetro “a” que representa o grau de discriminação do item está relacionado matematicamente à derivada da tangente da curva no ponto “b” (ponto de inflexão, ou seja, ponto onde a curva muda de concavidade) o que pode ser verificado graficamente com a inclinação da curva característica do item nas Figuras 4.1 e 4.2, verifica-se que o item 40, com $a=1,530$ apresenta uma inclinação superior a do item 26 que possui um valor de $a=0.972$, isto quer dizer que o item 40 tem uma maior facilidade de discriminar *sites* com usabilidades semelhantes, em torno do valor b , do que o item 26, evidentemente que cada um em sua área de maior informação.

4.3 Estimação das Habilidades

A partir dos resultados da fase anterior, faz-se nesta fase, a estimação da habilidade dos respondentes, neste caso a usabilidade dos *sites*. Como os itens são estimados na escala de usabilidade, é possível identificar onde o instrumento (conjunto de itens) contém maior informação. A Figura 4.3 representa a curva característica do item 42 (à esquerda) e a curva de informação do mesmo item (à direita). Pode-se perceber que este item é bastante útil para avaliar *sites* que tenham usabilidade em torno de -1,340 na escala gerada, e não apresenta informação considerável para *sites* que apresentem usabilidade maior que 1.

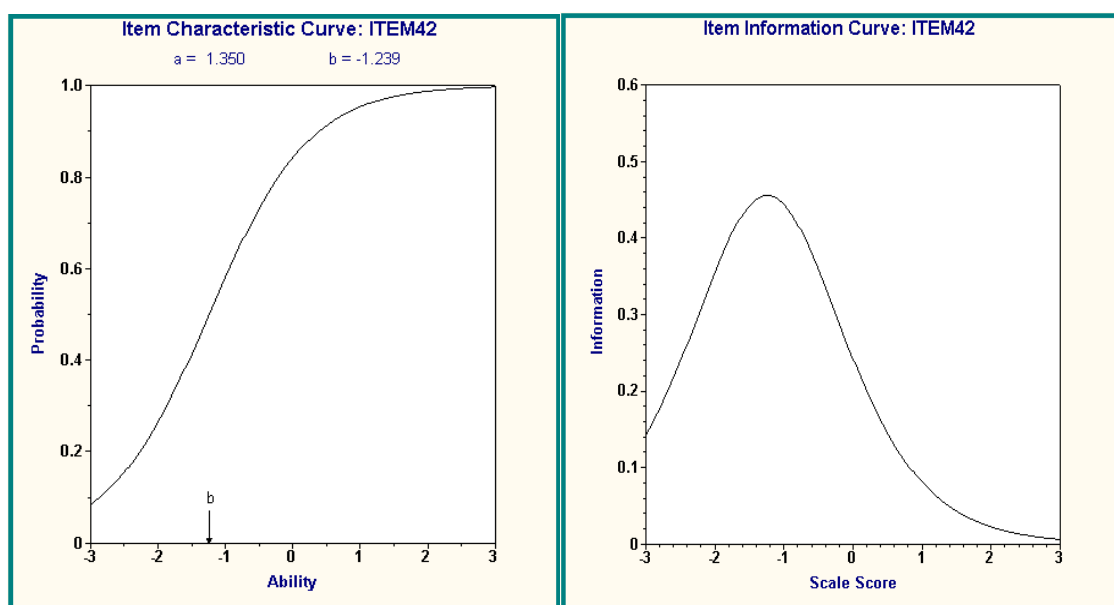


Figura 4.3 - Curva Característica e curva de informação do item 42

A curva característica de cada um dos 44 itens pode ser visualizada no Apêndice C, e a curva de informação no Apêndice D.

Nesta fase também é gerado o gráfico de informação total do conjunto de itens, que é a soma da informação de todos os itens. Este gráfico demonstra para qual região da escala gerada o construto (ou conjunto de itens) é mais indicado, ou seja, se ele é mais indicado para avaliar *sites* com baixa, média ou alta usabilidade. A forma desta curva indica a validade ou não do construto. A Figura 4.4 representa a curva de informação total do construto proposto. Verifica-se que existe uma boa definição da curva, o que valida o construto. E conclui-se que o construto é indicado para medir *sites* que apresentam baixa usabilidade, apresentando pouca informação para *sites* com usabilidade maior que 1,2 na escala proposta.

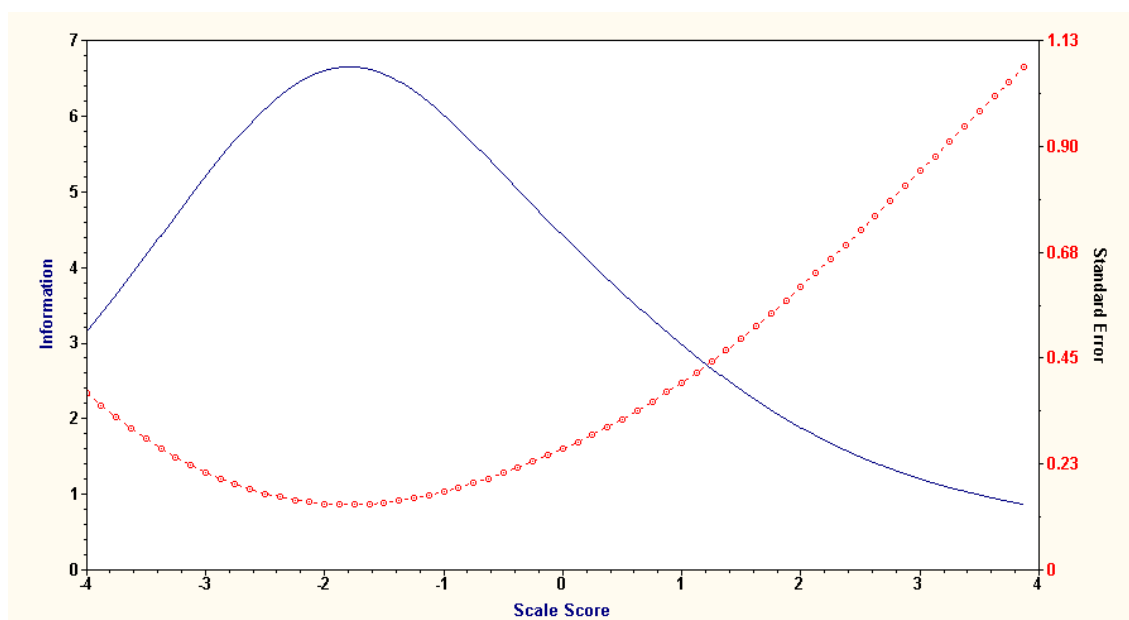


Figura 4.4 - Curva de informação total (Função de informação do teste – linha contínua; Erro padrão de medida – linha pontilhada)

A curva de informação total demonstra também o comportamento do erro padrão associado à precisão do instrumento. Verifica-se que a curva do erro-padrão atinge seu valor mínimo exatamente no ponto da escala onde a curva de informação atinge seu máximo. Desta forma, o instrumento é indicado para avaliar *sites* que estejam na região onde a curva de informação é maior que a curva do erro-padrão.

A média da usabilidade dos *sites* analisados, em relação à escala construída (média 0 e desvio padrão 1), foi de -0,0044 com desvio padrão igual a 0,8799. A Figura 4.5 demonstra a distribuição do grau de usabilidade dos *sites* utilizados para validação do construto.

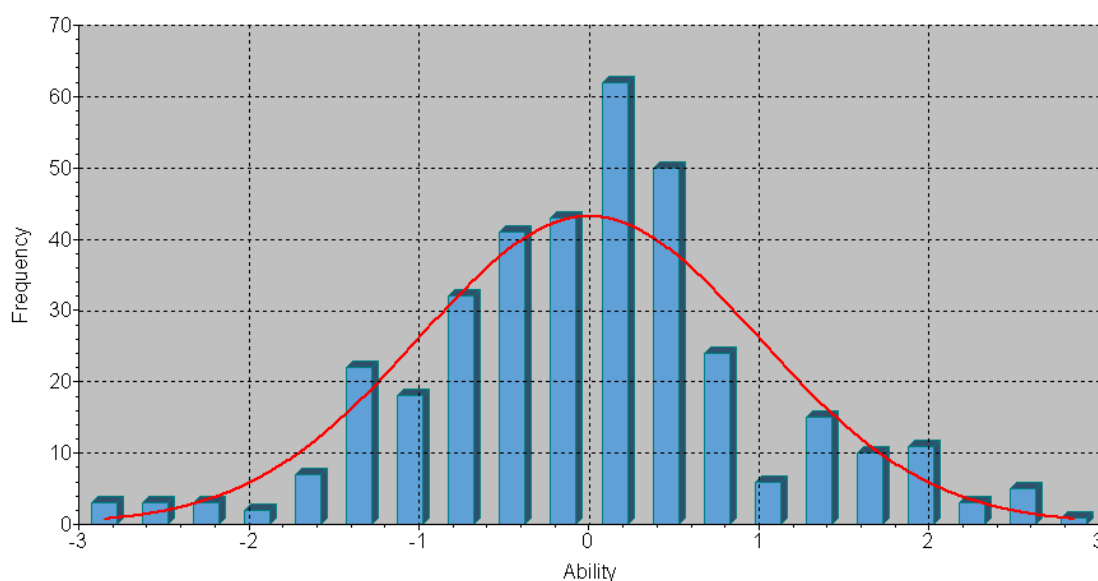


Figura 4.5 - Distribuição do grau de usabilidade dos *sites* utilizados

Verifica-se que a amostragem analisada possui um comportamento homogêneo em relação ao grau de usabilidade. Entretanto, verifica-se na Figura 4.4 que o instrumento criado não é o ideal para medir com precisão a amostragem, uma vez que grande parte de seu poder de discriminação está entre -4 e 1,2 enquanto a amostragem apresenta grau de usabilidade entre -2 e 2. Desta forma, teoricamente o que se esperava é que itens com parâmetro de dificuldade baixo apresentassem parâmetro de discriminação alto, o que pode ser confirmado no Quadro 4.2.

4.4 Construção da Escala de Medida do Grau de Usabilidade

Antes de partir para a construção final da escala, é necessário realizar uma análise aprofundada dos resultados obtidos nas estimações dos parâmetros. Esta análise consiste em identificar itens que não estejam alinhados com o construto e que portanto, não devem fazer parte do conjunto final. A identificação destes itens pode ser feita com base nos valores estimados para seus respectivos parâmetro de discriminação (a) e/ou parâmetro dificuldade (b).

Como discutido na seção 4.2, assume-se que um item com parâmetro de discriminação inferior a 0,700, na escala (0,1), possui baixo poder de discriminação, consequentemente

baixa informação, não sendo portanto adequado ao construto. Estes itens podem ser visualizados no Quadro 4.2.

O valor do parâmetro de dificuldade “b”, como discutido na seção 4.2, não indica a qualidade de um item, e sim a posição deste na escala construída. Entretanto, um conjunto de itens possui uma região em que sua precisão é maior, e esta coincide com a região onde os valores da função informação total são maiores que a função erro padrão total (Figura 4.4). Portanto, no presente construto é importante analisar a precisão de itens com parâmetro de dificuldade maior que 1,2 e menor que -4,2, uma vez que estes são os pontos representados na Figura 4.4 onde a função erro padrão total supera a função informação total.

Os itens com b maior que 1,2 foram: 03, 07, 12, 24, 26, 36 e 44. O item 30 apresentou b menor que -4,2. Dentre estes, os itens 12, 26, 30 e 44 apresentaram parâmetro de discriminação maior que 0,700, erro padrão e curva de informação aceitável quando comparado com os demais, o que motivou manter estes itens no construto, apesar de estarem, teoricamente segundo a curva de informação total, fora da área de precisão do instrumento. Os itens 03, 07, 24 e 36 apresentaram baixo poder de discriminação, e baixa informação o que reforça a inadequação destes ao construto proposto.

Sendo assim, decidiu-se pela eliminação apenas dos itens que apresentaram parâmetro de discriminação inferior a 0,700 que foram: 01, 03, 05, 07, 08, 09, 13, 24, 32, 33, 36, 37.

Com o objetivo de verificar se a retirada destes itens gera algum tipo de instabilidade no processo de estimação dos parâmetros e para confirmar a propriedade de interdependência dos itens realizou-se a retirada dos itens um a um em ordem crescente de valor do parâmetro de discriminação a . O resultado deste processo mostrou que não existe impacto negativo considerável na estimação dos parâmetros. O que se pode verificar foram alterações pequenas nos valores estimados e nos erros-padrão o que é inerente a um processo de interação deste porte. Todos os parâmetros estimados em cada uma dessas simulações podem ser visualizados no Apêndice E. O Quadro 4.3 traz os parâmetros estimados dos 32 itens restantes.

Item	Parâmetro a	Erro Padrão	Parâmetro b	Erro Padrão	Item	Parâmetro a	Erro Padrão	Parâmetro b	Erro Padrão
02	0,760	0,149	-1,541	0,302	25	0,975	0,17	-0,949	0,173
04	0,885	0,160	-1,463	0,254	26	1,086	0,193	1,682	0,257
06	0,824	0,194	-2,939	0,605	27	0,935	0,225	-2,845	0,562
10	0,787	0,148	-0,902	0,204	28	1,106	0,323	-2,866	0,610
11	0,798	0,148	-1,565	0,293	29	1,052	0,234	-2,149	0,368
12	0,781	0,217	4,685	1,195	30	0,752	0,231	-4,490	1,260
14	1,044	0,238	-2,298	0,400	31	0,872	0,181	-2,049	0,375
15	0,934	0,225	-2,920	0,579	34	1,090	0,174	-0,535	0,123
16	0,967	0,164	0,177	0,125	35	0,757	0,157	-1,145	0,245
17	1,424	0,402	-2,500	0,427	38	1,036	0,228	-2,959	0,528
18	0,768	0,178	-2,672	0,561	39	1,276	0,265	-1,637	0,244
19	1,293	0,232	0,452	0,113	40	1,544	0,311	-1,746	0,233
20	0,750	0,152	-0,282	0,167	41	1,277	0,293	-2,015	0,329
21	1,176	0,305	-2,544	0,467	42	1,437	0,227	-1,205	0,150
22	0,914	0,162	0,185	0,133	43	1,107	0,201	-1,694	0,255
23	1,152	0,246	-1,946	0,304	44	0,690	0,175	4,420	1,061

Quadro 4.3 - Estimação dos parâmetros com os itens seleccionados

Observa-se que na re-estimação dos parâmetros após a retirada dos itens, com baixo poder de discriminação, houve pequena variação nos valores, o que é normal num processo iterativo. Todavia, verifica-se particularmente o item 44 que antes possuía um parâmetro de discriminação igual a 0,716 agora passou a apresentar o mesmo parâmetro igual a 0,690, uma variação de 3% o que é tolerável no contexto, mas seguindo o critério teórico apresentado na seção 4.2 este item passa a apresentar poder de discriminação abaixo de 0,700. Desta forma, estatisticamente este item estaria eliminado do construto. Entretanto, do ponto de vista prático isto não significa que este item não esteja medindo grau de usabilidade. Apesar deste item estar fora da região de precisão do instrumento seu poder de discriminação foi mais adequado que os demais itens eliminados. Sendo assim, manteve-se este item para a próxima fase da construção da escala, que foi a definição dos itens e níveis âncora. A curva de informação total também sofreu pouca variação, e pode ser comparada com a curva de informação total dos 44 itens antes da eliminação no Apêndice F.

4.4.1 Itens eliminados

Até este ponto, o construto foi avaliado exclusivamente do ponto de vista estatístico, entretanto, é importante ressaltar que a construção de uma escala de medida de usabilidade,

ou qualquer outra escala prática, necessita de reflexões práticas referente ao assunto que se está propondo medir. Desta forma, discute-se aqui a eliminação dos itens sob o ponto de vista prático.

Os itens eliminados, juntamente com suas descrições podem ser verificados no Quadro 4.4.

Item	Descrição
01	Ao abrir a página inicial, não há abertura de janelas pop-up?
03	A página inicial possui menos de 2,3 páginas de rolagem?
05	O layout da página é adaptável ao tamanho da janela, pelo menos horizontalmente (ou seja não há rolagem horizontal)?
07	O site possui a opção “ajuda”
08	Na área principal do site, existem textos com no máximo quatro cores diferentes?
09	Na área principal do site, existem textos com no máximo três fontes diferentes?
13	É possível encontrar informações sobre a empresa (endereço físico, telefone etc)?
24	O tamanho da fonte dos textos é de dez pontos ou mais?
32	Os campos destinados a formulários de preenchimento pelo usuário estão agrupados linearmente, evitando espaços desnecessários?
33	No preenchimento de um formulário, as opções que não são válidas ou não estão disponíveis estão visivelmente desativadas, prevenindo erros?
36	É possível efetivar uma compra sem realizar um cadastro (que inclua login e senha)?
37	O sistema emite sinais sonoros quando ocorrem problemas na entrada de dados?

Quadro 4.4 - Descrição dos itens eliminados

As recomendações pesquisadas são unânimes em considerar a característica levantada no item 01 como muito importante para a usabilidade de um site. Entretanto, a presente pesquisa identificou que não se trata de uma característica cumulativa, ou seja, não é verdade afirmar que quanto menor o grau de usabilidade de um site, maior será a probabilidade deste possuir *pop-up*. Isto se dá possivelmente pelo fato de que para um site fazer uso desta “ferramenta” o mesmo deve possuir certa maturidade técnica e determinados interesses comerciais, não sendo, portanto o modelo logístico de dois parâmetros a ferramenta ideal para modelar esta característica.

Os itens 03 e 05 referem-se ao elemento de design – rolagem. E conforme discutido na seção 3.1.1.1 a utilização ou não de rolagem em uma interface gráfica não é um consenso na literatura. Existem autores que defendem a utilização de rolagem (BYRNE, ET. AL., 1999;

CAMPBELL; MAGLIO, 1999), outros que defendem a não utilização (NIELSEN, 1997; PIOLAT; ROUSSEY; THUNIN, 1998) e existem também os que defendem a utilização limitada de rolagem (DETWEILER; OMANSON, 1996; NIELSE, 1996; SPOOL, *ET AL*, 1997; SPYRIDAKIS, 2000; LYNCH; HORTON, 2002; NIELSEN; LORANGER, 2006). Portanto em termos práticos, a eliminação destes itens reflete principalmente esta falta de consenso indicando para estudos mais aprofundados no campo de usabilidade de interfaces, ou na aplicação de outros modelos estatísticos para modelar este comportamento.

O item 07 aborda o tipo de característica que segundo Cybis (2007) deve estar presente em todas as interfaces – a ajuda contextual e específica. A opção de ajuda pode auxiliar consideravelmente usuários novatos, e significar um acréscimo positivo na experiência do usuário com o site (HHS, 2006). Entretanto, esta característica está longe de ser básica, o que é comprovado com o valor estimado do parâmetro de dificuldade acima da média ($b=1,798$). A eliminação deste item, não significa que ele não represente uma característica importante para a usabilidade, todavia, não se apresentou alinhado com o construto desenvolvido. Isto pode ser justificado pelo fato de que a opção “ajuda” não tem função direta na tarefa de compra em um site comercial, ela é utilizada principalmente quando alguma das características básicas de usabilidade falha. Portanto, a análise estatística demonstra que este item não é indicado para o presente construto, porém, a prática indica que esta característica está relacionada com um contexto mais amplo de usabilidade merecendo assim estudos posteriores, talvez em outro construto mais amplo.

Os itens 08 e 09 estão ligados ao aspecto visual e a legibilidade de um site. Foram elaborados com base em Nielsen e Loranger (2006) onde os autores recomendam no máximo quatro cores diferentes e três fontes nas principais áreas do site, caso contrário ele parecerá desestruturado e amador. Entretanto, estes números não são unanimidade na literatura. HHS (2006) afirma que os usuários podem distinguir até 10 cores diferentes e fazer relações informativas entre elas, no entanto recomenda a utilização de no máximo cinco cores diferentes. Outros autores, como Detweiler; Omanson, (1996); Nygren; Allard, (1996) Spyridakis, (2000) e Chaparro; Bernard, (2001) defendem que as cores e as fontes diferenciadas numa página devem permitir a fácil identificação e entendimento das informações contidas, não citando quantidades específicas. Sendo assim, unindo estes argumentos e a análise estatística desenvolvida no presente trabalho pode-se concluir que em termos de usabilidade, a quantidade específica de no máximo quatro cores diferentes, ou 3 fontes diferentes não indicam necessariamente baixa usabilidade, dependendo, portanto,

segundo a teoria sobre usabilidade, de outros fatores tais como a organização destas cores e fontes na página e seus significados no contexto.

O item 13 está relacionado a possibilidade de encontrar informações sobre a empresa de *e-commerce*. Foi construído baseado na recomendação de Nielsen e Loranger (2006) em que os autores argumentam que em uma compra on-line os usuários necessitam estabelecer uma relação de confiança e credibilidade com o site e na maioria das vezes eles procuram informações sobre a empresa (endereço físico, histórico, telefone, etc) e ficam desorientados quando não encontram. Porém, na prática esta característica parece mais relacionada a confiabilidade de um site do que com sua facilidade de uso, ou seja, se o usuário não encontrar as informações, embora elas existam em alguma parte do site, eles consideram um problema de confiabilidade e não de usabilidade, o que foi confirmado pelo modelo estatístico da TRI.

O item 24 diz respeito ao tamanho da fonte e estipula um tamanho maior ou igual a 10 pontos que é o recomendado por Nielsen e Loranger (2006). HHS (2006) recomenda fonte de 12 pontos para facilitar a leitura, e reforçam que jamais se utilize fonte inferior a 9 pontos. Bernard; Liao e Mills (2001) e Ellis e Kurniawan (2000) recomendam 14 pontos, principalmente pelo fato de facilitar a leitura de usuários com alguma deficiência visual. Sendo assim, não parece haver na literatura um consenso quanto ao tamanho mínimo ideal para texto na web. Isto se dá pelo fato de depender de alguns fatores externos tais como, o tipo de navegador utilizado, o público-alvo e demais fatores ligados a legibilidade (contraste com o fundo, densidade do texto etc). Através da análise teórica e prática, e da análise estatística verifica-se que esta não é uma característica que possa ser definida por apenas um tamanho de fonte dependendo de outros fatores. Principalmente em *sites de e-commerce*, onde a tendência é apresentar textos breves.

O item 32 é baseado em HHS (2006) onde é recomendado que em formulários de preenchimento pelo usuário, os rótulos dos campos de entrada de dados estejam agrupados de forma a facilitar o preenchimento. Este agrupamento tem a função de auxiliar no entendimento do relacionamento dos dados com seus rótulos. Apesar de existir argumentos teóricos e práticos indicando que esta característica contribui para a usabilidade de uma interface gráfica, o item não se apresentou estatisticamente adequado para este construto. Assim como os itens 33 e 36 que também avaliam características relacionadas ao preenchimento de formulário. O que leva a conclusão de que o construto avalia de forma mais precisa as características relacionadas ao design da página, a navegação, a ferramenta de

busca e seus resultados. Indicando que a avaliação de aspectos relacionados à efetivação da compra propriamente dita necessita de um construto mais amplo com outros itens.

Teoricamente o item 37 tende a contribuir à usabilidade de uma interface, entretanto, na prática, sua efetividade depende de outros fatores, tais como o contexto de navegação e a necessidade de ativação da função de som, quando existir. O contexto de navegação em que um aviso sonoro seria importante é aquele em que “a tarefa force o usuário a olhar fora da tela” (CYBIS 2007 p.54), o que na maioria das vezes não é o caso da tarefa de compra em *sites* de *e-commerce*. Desta forma, pode-se concluir que esta dependência de outros fatores também contribuiu para a eliminação estatística do item 37.

4.4.2 Mudança de escala

Uma vez que, tanto os parâmetros dos itens quanto as proficiências dos *sites* são estimadas na mesma métrica, é possível criar uma escala padrão para medição do grau de usabilidade baseado nos itens elaborados. Esta escala pode manter a métrica gerada, ou seja, média 0 (zero) e desvio padrão 1 (um) ou criar uma nova métrica através de uma transformação linear, que altera apenas a representação da escala, mantendo as relações de ordem existentes. Para facilitar a visualização da escala e a definição dos itens âncoras, neste trabalho empregou-se a métrica (50, 10), ou seja, média 50 e desvio padrão 10. Para esta transformação utilizou-se as equações (4.1), (4.2), (4.3) e (4.4), que transformam os parâmetros (a,b) e a proficiência (θ) da antiga escala (0,1) para a nova escala (50,10).

$$\theta^* = 10 \times \theta + 50 \quad (4.1)$$

$$b^* = 10 \times b + 50 \quad (4.2)$$

$$a^* = \frac{a}{10} \quad (4.3)$$

$$P(U_i = 1 | \theta) = P(U_i = 1 | \theta^*) \quad (4.4)$$

A escala gerada seja ela na métrica (0,1), (50,10) ou qualquer outra, é a princípio de natureza arbitrária, necessitando de uma interpretação que realize a ligação dos valores estimados com o conteúdo existente nos itens. O procedimento utilizado para realizar esta interpretação foi a identificação dos níveis âncora e de itens âncora.

4.4.3 Níveis âncora e itens âncora para a escala de usabilidade

Níveis âncora são pontos selecionados pelo analista na escala da habilidade para serem interpretados na prática. Já os itens âncora são itens selecionados, segundo a definição dada nas equações (4.5), (4.6) e (4.7), para cada um dos níveis âncora (ANDRADE, TAVARES e VALLE, 2000).

$$P(U = 1 | \theta = Y) \geq 0,65 \quad (4.5)$$

$$P(U = 1 | \theta = X) \leq 0,50 \quad (4.6)$$

$$P(U = 1 | \theta = Y) - P(U = 1 | \theta = X) \geq 0,30 \quad (4.7)$$

O que significa que para um item ser considerado âncora em um determinado nível âncora da escala de grau de usabilidade, ele precisa ser respondido positivamente por pelo menos 65% dos *sites* com este nível de usabilidade e por uma proporção menor de 50% de *sites* com o nível de usabilidade imediatamente inferior, sendo que a diferença entre a proporção de *sites* destes dois níveis de usabilidade deve ser de pelo menos 30%. Assim, um item é considerado âncora quando for típico daquele nível âncora, ou seja, o requisito contido neste item é bastante observado nos *sites* com grau de usabilidade Y e pouco observado em *sites* com grau de usabilidade X.

Para um item ser âncora, ele precisa satisfazer as três condições representadas pelas equações (4.5), (4.6) e (4.7) simultaneamente. É comum encontrar-se estudos em que itens que atendam a pelo menos duas das condições também sejam considerados, definidos assim como “quase âncora” (VARGAS, 2007). Neste estudo, também foram considerados itens “quase âncora”.

Baseado nesta definição e nas condições de níveis âncora, realizou-se a identificação destes na escala gerada com as novas estimativas, agora na métrica (50, 10) que pode ser visualizada no Quadro 4.5. Verificam-se quatro níveis âncoras e “quase âncora”, seis itens âncora (com proficiências destacadas em cinza escuro) e 14 itens “quase âncora” (com proficiência destacadas em cinza claro). Os níveis 0, 10, 20, 70, 80, 90 e 100, segundo as condições adotadas, não identificaram nenhum item âncora ou “quase âncora”.

Item	Parâmetro		Grau de Usabilidade (Θ) (50,10)											P(Z)-P(Y)
	a	b	0	10	20	30	40	50	60	70	80	90	100	
02	0,08	34,59	0,07	0,13	0,25	0,41	0,60	0,76	0,87	0,94	0,97	0,99	0,99	0,19
04	0,09	35,37	0,04	0,10	0,20	0,38	0,60	0,78	0,90	0,96	0,98	0,99	1,00	0,22
06	0,08	20,61	0,15	0,29	0,49	0,68	0,83	0,92	0,96	0,98	0,99	1,00	1,00	0,19
10	0,08	40,98	0,04	0,08	0,16	0,30	0,48	0,67	0,82	0,91	0,96	0,98	0,99	0,19
11	0,08	34,35	0,06	0,13	0,24	0,41	0,61	0,78	0,89	0,95	0,97	0,99	0,99	0,20
12	0,08	96,85	0,00	0,00	0,00	0,01	0,01	0,03	0,05	0,11	0,21	0,37	0,56	0,19
14	0,10	27,02	0,06	0,14	0,32	0,58	0,79	0,92	0,97	0,99	1,00	1,00	1,00	0,26
15	0,09	20,80	0,13	0,27	0,48	0,70	0,86	0,94	0,97	0,99	1,00	1,00	1,00	0,22
16	0,10	51,77	0,01	0,02	0,04	0,11	0,24	0,46	0,69	0,85	0,94	0,98	0,99	0,23
17	0,14	25,00	0,03	0,11	0,33	0,67	0,89	0,97	0,99	1,00	1,00	1,00	1,00	0,34
18	0,08	23,28	0,14	0,27	0,44	0,63	0,78	0,89	0,94	0,97	0,99	0,99	1,00	0,19
19	0,13	54,52	0,00	0,00	0,01	0,04	0,13	0,36	0,67	0,88	0,96	0,99	1,00	0,31
20	0,08	47,18	0,03	0,06	0,12	0,22	0,37	0,55	0,72	0,85	0,92	0,96	0,98	0,18
21	0,12	24,56	0,05	0,15	0,36	0,66	0,86	0,95	0,98	1,00	1,00	1,00	1,00	0,30
22	0,09	51,85	0,01	0,02	0,05	0,12	0,25	0,46	0,68	0,84	0,93	0,97	0,99	0,22
23	0,12	30,54	0,03	0,09	0,23	0,48	0,75	0,90	0,97	0,99	1,00	1,00	1,00	0,27
25	0,10	40,51	0,02	0,05	0,12	0,26	0,49	0,72	0,87	0,95	0,98	0,99	1,00	0,23
26	0,11	66,82	0,00	0,00	0,01	0,02	0,05	0,14	0,32	0,59	0,81	0,93	0,97	0,27
27	0,09	21,55	0,12	0,25	0,46	0,69	0,85	0,93	0,97	0,99	1,00	1,00	1,00	0,23
28	0,11	21,34	0,09	0,22	0,46	0,72	0,89	0,96	0,99	1,00	1,00	1,00	1,00	0,26
29	0,11	28,51	0,05	0,12	0,29	0,54	0,77	0,91	0,96	0,99	1,00	1,00	1,00	0,25
30	0,08	5,10	0,41	0,59	0,75	0,87	0,93	0,97	0,98	0,99	1,00	1,00	1,00	0,18
31	0,09	29,51	0,07	0,15	0,30	0,50	0,71	0,86	0,93	0,97	0,99	0,99	1,00	0,21
34	0,11	44,65	0,01	0,02	0,06	0,17	0,38	0,64	0,84	0,94	0,98	0,99	1,00	0,26
35	0,08	38,55	0,05	0,10	0,20	0,34	0,53	0,70	0,84	0,92	0,96	0,98	0,99	0,19
38	0,10	20,41	0,11	0,25	0,49	0,73	0,88	0,96	0,98	0,99	1,00	1,00	1,00	0,24
39	0,13	33,63	0,01	0,05	0,15	0,39	0,69	0,89	0,97	0,99	1,00	1,00	1,00	0,30
40	0,15	32,54	0,01	0,03	0,13	0,40	0,76	0,94	0,99	1,00	1,00	1,00	1,00	0,36
41	0,13	29,85	0,02	0,07	0,22	0,50	0,79	0,93	0,98	0,99	1,00	1,00	1,00	0,29
42	0,14	37,95	0,00	0,02	0,07	0,24	0,57	0,85	0,96	0,99	1,00	1,00	1,00	0,33
43	0,11	33,06	0,03	0,07	0,19	0,42	0,68	0,87	0,95	0,98	0,99	1,00	1,00	0,26
44	0,07	94,20	0,00	0,00	0,01	0,01	0,02	0,05	0,09	0,16	0,27	0,43	0,60	0,17

Quadro 4.5 - Classificação dos itens âncora e “quase âncora” de acordo com os níveis do grau de usabilidade – na métrica (50,10).

A identificação dos níveis âncora e dos itens âncora indica qual a região da escala gerada onde há maior informação, e a partir daí, é possível identificar o requisito de usabilidade predominante em cada um destes níveis. Desta forma, pode-se construir uma escala onde se identifique quais requisitos correspondem ao maior ou menor grau de usabilidade. No Quadro 4.6, foram descritos e classificados os itens referentes ao nível 30 de grau de usabilidade.

Nível 30		
ITEM	DESCRIÇÃO	CONCEITO DE USABILIDADE ASSOCIADO
06	As opções principais do site estão visíveis? Ou seja, não há necessidade de passar o mouse sobre uma área gráfica para ver opções ocultas?	Design da página
15	Os parágrafos de texto são separados?	Design da página
17	Os títulos de telas, janelas e caixas de diálogo estão no alto, centrados ou justificados à esquerda?	Design da página
21	O preço de um produto consta ao lado da imagem ou do link do produto?	Busca (informações)
27	Os grupos de botões de comando estão dispostos em coluna e à direita, ou em linha a abaixo dos objetos aos quais estão associados?	Design da página
28	Quando há rolagem, não existem elementos de design (na tela inicial) que pareçam com marcadores de final de página?	Design da página
38	As mensagens de erro estão isentas de abreviaturas e/ ou códigos gerados pelo sistema operacional?	Navegação

Quadro 4.6 - Descrição e classificação dos itens posicionados no nível 30 na escala gerada, de acordo com os aspectos de usabilidade associados.

Considerando-se que o nível 30 possui, na escala proposta, um grau de exigência de usabilidade baixo, os itens âncora localizados neste nível, consequentemente, abordam requisitos básicos de usabilidade. Sendo assim, *sites* que responderem positivamente os itens âncora posicionados neste nível têm no mínimo 0,65 de probabilidade de possuir grau de usabilidade maior ou igual a 30. E *sites* que não apresentarem os requisitos questionados nestes itens têm probabilidade inferior a 0,65 de estarem neste nível. Observa-se que o conceito de usabilidade mais presente neste nível é o referente ao design da página, ligado à exigência de memorização, legibilidade, padrões da web e modelos mentais.

Na prática, o item 06 está associado à carga de memorização imposta ao usuário, ou seja, não manter as opções visíveis durante as ações do usuário requer que ele memorize as opções ocultas ou retorne a elas sempre que ficar em dúvida, o que prejudica substancialmente a usabilidade de um site. O item 15 está ligado à legibilidade de um texto qualquer. A utilização de espaços entre parágrafos é uma característica simples que ajuda a tornar o texto mais agradável e é considerada uma característica básica, mas muito importante na web, onde a leitura é menos confortável que nos demais meios (HFDG, 2001). Os itens 17, 27 e 28 estão relacionados a padrões da web que por sua vez estão associados aos modelos mentais do

usuário, ou seja, definem características básicas que um usuário espera de um website. Segundo HHS (2006) usuários podem criar expectativas baseadas em seus conhecimentos prévios ou experiências passadas. A importância destes padrões é reforçada pelo fato dos usuários gastarem a maior parte do tempo em outros *websites* (NIELSEN; TAHIR, 2002). Mesmo se multidões de usuários visitarem determinado site porque ele é o maior e o mais proeminente na Web, as visitas acumuladas a outros *sites* continuarão a superar muito o número de visitas a este. Isso significa que os usuários criam expectativas em relação a um site de acordo com aquilo que eles aprenderam a esperar de outros. Se eles estiverem acostumados a padrões e convenções de design predominantes, vão esperar encontrá-los também em qualquer site (NIELSEN; LORANGER, 2006).

O item 21 levanta uma característica primária na busca por qualquer produto – o custo. Através de pesquisa com usuários de *e-commerce*, Nielsen e Loranger (2006) concluíram que uma das primeiras coisas que os usuários querem saber sobre um produto ou serviço é o preço. Segundo os autores um preço pode informar algumas coisas de uma vez:

- o valor do produto;
- se o produto está ou não dentro de seu orçamento;
- se o cliente está comprando no segmento certo.

Sendo assim, dificultar o acesso do usuário a esta informação pode significar insatisfação para o usuário e insucesso para o site.

Quando ocorrem erros inevitáveis na navegação, uma das primeiras coisas que o sistema necessita apresentar é um claro diagnóstico do problema ao usuário. “Caso a correção do erro seja efetuada pelo sistema, a interface deve apresentar uma mensagem de confirmação que indique ao usuário a estratégia de recuperação planejada” (CYBIS, 2007 p53). Este tipo de feedback básico que pode facilitar a interação usuário-interface, é tratado o item 38.

Observa-se que, na prática, todas estas características são consideradas básicas e observadas na maioria dos *sites* de *e-commerce*, o que possibilitou o posicionamento destes itens no nível 30, convergindo a prática com os resultados estatísticos. Desta forma, *sites* que não apresentem estas características têm grande chance de estar abaixo do nível 30 na escala de grau de usabilidade gerada.

Na escala proposta, considera-se o grau de usabilidade de nível 40 maior que o nível 30, sendo assim, os itens representados por aquele nível possui maior exigência que os itens representados pelo nível 30. Sendo assim, *sites* que apresentarem os requisitos abordados nos

itens âncora do nível 40 (Quadro 4.7) têm também 0,65 de probabilidade de possuir grau de usabilidade maior ou igual a 40.

Nível 40		
ITEM	DESCRIÇÃO	CONCEITO DE USABILIDADE ASSOCIADO
23	A maioria dos produtos possui informações sobre eles?	Busca (informações)
31	É possível saber os custos totais antes de fazer cadastro?	Busca (informações)
39	Todas as páginas possuem os mesmos <i>layouts</i> e exibem ao usuário as mesmas características?	Navegação
40	O logotipo da empresa está no canto superior esquerdo em todas as páginas do site?	Navegação
41	Existe um link de um único clique que conduz a página inicial?	Navegação
42	Qualquer ação do usuário pode ser revertida através da opção “desfazer” ?	Navegação
43	O site permite navegação em suas páginas em apenas uma janela, ou seja, não há abertura de novas janelas em meio a navegação?	Navegação

Quadro 4.7 - Descrição e classificação dos itens posicionados no nível 40 na escala gerada, de acordo com os aspectos de usabilidade associados.

Sendo que o construto proposto é cumulativo, *sites* que apresentarem as características abordadas pelos itens âncora de nível 40 provavelmente apresentaram também as características abordadas pelos itens âncora de nível 30. Esta afirmativa pode ser comprovada através da comparação prática dos itens. A maioria dos itens pertencentes ao nível 40 aborda características de navegação, que estão ligadas diretamente com a interação ativa entre usuário-site. Enquanto que as características predominantes no nível 30 estão relacionadas com o design da página que está mais ligada a interação visual. Falha nas características visuais pode confundir o usuário e tornar a interação mais difícil, entretanto, não necessariamente impede o usuário de realizar determinada ação. Já problemas ligados a navegação podem impedir, ou dificultar substancialmente a realização de determinada ação. Sendo assim, na prática pode-se afirmar que características relacionadas a navegação agregam maior grau de usabilidade que características associadas ao design da página, o que foi comprovado pela análise estatística. Por exemplo, apresentar os títulos de janelas e caixas de diálogo no alto, centrados ou justificados à esquerda (item 17) pode facilitar a visualização destes, mas dificilmente impedirão o usuário de vê-los. Já apresentar layout e características distintas entre a página principal e as páginas internas (item 39) pode desconcentrar o usuário

e impedi-lo de fazer associações informativas importantes durante uma compra on-line (NIELSEN, 1993; NIELSEN; LORANGER, 2006). Sendo assim, é compreensível que o item 39 possua grau de usabilidade maior que o item 17. O mesmo comportamento pode ser verificado nos demais itens do nível 30 e 40.

No entanto, algumas características de design da página estão associadas, na prática, não só a aspectos visuais, mas também ao desempenho da tarefa, como é o caso do item 10, pertencente, segundo a análise, ao nível ancora 50 (Quadro 4.8).

Combinar lista e tarefas em uma ordem melhora a eficiência e o desempenho de usuários em um site (HHS 2006). Desta forma, é compreensível que características como esta assumam um grau de usabilidade superior, uma vez que associa design, navegação e eficiência, como é o caso da disposição dos objetos de interação de uma caixa de diálogo em uma ordem lógica (item 10). O item 25 trás uma característica muito importante em uma experiência de compra em *sites* de *e-commerce*: a visualização detalhada do produto. Segundo Nielsen e Loranger (2006) o pior erro é quando um usuário clica no botão “Ampliar foto” e o site exibe a mesma foto. Esses links e botões que não funcionam desperdiçam tempo e confundem ainda mais o usuário. Ver um detalhe específico ou avaliar uma textura pode dar ao comprador a confiança de que precisam para tomar uma decisão de compra. A análise posicionou esta característica no nível ancora 50 e a prática a classifica como um diferencial de usabilidade capaz de facilitar e influenciar a decisão de compra o que afirmar que existe convergência entre a análise e a prática. O mesmo ocorre com o item 34 que também representa um diferencial de usabilidade, agora focado na redução de erros de entrada em formulários.

Nível 50		
ITEM	DESCRIÇÃO	CONCEITO DE USABILIDADE ASSOCIADO
10	A disposição dos objetos de interação de uma caixa de diálogo segue uma ordem lógica?	Design da página
25	É possível ampliar as fotos dos produtos para visualizar detalhes?	Busca (informações)
34	No preenchimento de um formulário, é informada a forma de preenchimento?	Suporte à tarefa de compra

Quadro 4.8 - Descrição e classificação dos itens posicionados no nível 50 na escala gerada, de acordo com os aspectos de usabilidade associados.

Dentro do construto proposto, o nível 60 foi o maior nível que apresentou itens âncora (Quadro 4.9). Desta forma, *sites* que apresentarem as características abordadas por estes itens

possivelmente apresentarão também as características abordadas nos itens dos níveis anteriores. Este nível se caracterizou, assim como o nível 50 por itens mais elaborados e que abordam diferenciais de usabilidade, agora relacionados com a navegação, como é o caso dos itens 16 e 22. Dentre os padrões utilizados na web existem alguns associados à clicabilidade como, por exemplo, palavras coloridas, sublinhadas ou em cor azul dão uma forte percepção de clicabilidade (NIELSEN; LORANGER, 2006). Esta é uma padronização, abordada no item 16, que facilita a navegação e que poucos *sites* aplicam confundindo alguns usuários. O item 22 diz respeito à orientação quanto à navegação na arquitetura de informação em um site. O item 19 é uma característica que representa um diferencial no tratamento do resultado da busca ou seleção de produtos.

Nível 60		
ITEM	DESCRIÇÃO	CONCEITO DE USABILIDADE ASSOCIADO
16	As palavras aparentemente clicáveis são de fato clicáveis ?	Navegação
19	Os resultados de busca ou seleção permitem classificação por outros critérios além de custo?	Busca (informações)
22	Existe uma orientação ao usuário quanto ao restante do site?	Navegação

Quadro 4.9 - Descrição e classificação dos itens posicionados de nível 60 na escala gerada, de acordo com os aspectos de usabilidade associados.

O Quadro 4.10 demonstra a distribuição dos itens de acordo com o conceito de usabilidade associado e os níveis âncora gerados pela análise. Observa-se que o nível 30 abrange principalmente os itens relacionados a conceitos de design da página enquanto o nível 40 abrange principalmente aspectos relacionados com a navegabilidade. O nível 50 aborda os conceitos relacionados à busca e suporte à tarefa de compras, bem como o item 10 que aborda design da página de uma forma mais elaborada que os itens pertencentes ao nível 30, como foi discutido anteriormente. No nível 60, a navegabilidade volta a ser o aspecto principal. Esta distribuição permite concluir que os aspectos relacionados a navegabilidade, quando comparados com os aspectos de design da página, exigem uma grau de usabilidade maior o que é confirmado pela literatura e pela prática em usabilidade, como discutido anteriormente. Sendo assim, este respaldo teórico e prático valida o construto proposto, permitindo a criação de uma escala do grau de usabilidade baseada nos itens elaborados.

CONCEITO DE USABILIDADE ASSOCIADO	GRAU DE USABILIDADE			
	30	40	50	60
Design da página	06, 15, 17, 27, 28	-	10	-
Navegação	38	39, 40, 41, 42, 43	-	16, 22
Busca (informações)	21	23, 31	25	22
Suporte à tarefa de compra	-	-	34	-

Quadro 4.10 - Distribuição dos itens de acordo com o conceito de usabilidade e os níveis na escala gerada.

Alguns itens não foram classificados nas condições impostas para representar item âncora, ou “quase âncora” porém, seus resultados podem ser avaliados sob o ponto de vista prático, deixando a cargo do analista de usabilidade decidir pela utilização ou não destes itens. Como é o caso dos itens 04, 12, 14, 18, 26, 30 e 44. Segundo a análise, o item 04 apresenta 0,60 de probabilidade de pertencer ao nível âncora 40 e se distancia 0,22 do nível anterior que apresenta 0,38 de probabilidade. A característica abordada neste item está relacionada à navegabilidade - As imagens, botões ou palavras clicáveis apresentam forma diferenciada quando são selecionadas? – Esta é uma característica importante para dar ao usuário noção de controle e feedback a suas ações. Tendo em vista as características abordadas nos demais itens pertencentes ao nível 40, pode-se recomendar o posicionamento do item 04 também no nível 40.

O item 12, juntamente com o item 44, representa os itens que abordam as características de maior grau de usabilidade. Apesar de não ter satisfeito as condições definidas para item âncora ou “quase âncora” pode-se afirmar que o item 12 tem 0,56 de probabilidade acumulada de estar no nível 100 e diferença de 0,19 em relação ao nível anterior. Enquanto o item 44 possui 0,60 de probabilidade acumulada de também estar no nível 100, e diferença de 0,17 em relação ao nível 90. Como pode ser visto no Quadro 4.3 ambos os itens apresentaram baixo parâmetro de discriminação o que pode justificar esta baixa diferença entre os níveis 90 e 100 o que dificulta a exatidão no posicionamento destes itens na escala. Entretanto, é possível afirmar que estes itens possuem grau de usabilidade maior que 60 (último nível âncora definido na escala). Do ponto de vista prático, o item 12, baseado em Nielsen e Norman (2001) representa uma opção diferenciada de navegação e é definido como: O site possui opção de acesso em outras línguas?. O Item 44 (Os links já visitados mudam de cor?) representa na prática, segundo Nielsen e Loranger (2006), um problema de usabilidade de alto

impacto na navegabilidade. “Não alterar as cores dos links cria confusão navegacional, pois os usuários não conseguem perceber claramente suas diferentes escolhas ou onde eles estão” (NIELSEN; LORANGER, 2006 p.61). A análise estatística mostrou que esta é uma das recomendações de usabilidade mais desrespeitadas na web, uma vez que poucos *sites* a apresentaram. Sendo assim, os resultados obtidos através da TRI juntamente com a análise sob a óptica da usabilidade permite concluir que este item possui um grau de usabilidade superior a 60, embora a distinção em qual nível exato pertença não fique clara. O item 14 trata basicamente de padrões visuais e padrões de legibilidade, semelhante aos itens 15, 17 e 27 que pertencem ao nível ancora 30. O resultado da análise do item 14 o deu uma probabilidade de 0,58 de pertencer ao nível 30 o que somado as características do conjunto de itens deste nível pode indicar que este item representa grau de usabilidade em torno de 30 também. O mesmo pode ser verificado no item 18, que apresentou 0,63 de probabilidade acumulada de pertencer ao nível 30.

O item 26 aborda a seguinte questão: Em produtos em que existam mais de uma perspectiva, é possível visualizar todas estas perspectivas? Na prática esta é uma característica que auxilia o usuário na visualização de detalhes de um produto evitando que ele procure por estes detalhes em fotos pouco informativas ou em outros lugares do site. É um item que vem a acrescentar ao item 25 (É possível ampliar as fotos dos produtos para visualizar detalhes?), posicionado no nível âncora 50. Sendo assim, é de se esperar que apresente grau de usabilidade igual ou superior a este. Isto foi confirmado pela análise, apesar de não ter sido classificado como item ancora, o item 26 apresentou uma probabilidade de 0,59 de pertencer ao nível 70 e 0,32 de pertencer ao nível 60.

O item 30 foi o item que apresentou menor grau de usabilidade, 0,59 de probabilidade de estar no nível 10 e 0,18 de diferença do nível 0 (zero). Representando assim uma característica com baixo grau de usabilidade podendo ser considerada abaixo do primeiro nível ancora definido na escala proposta que é 30.

4.4.4 Localização dos *sites* analisados na escala gerada

As estimações dos itens foi feita na mesma métrica da escala de usabilidade, o que permite a localização de cada um dos *sites* nos níveis de usabilidade determinados pelos itens. A Figura 8 ilustra a distribuição dos *sites* analisados em cada um dos níveis da escala de grau de usabilidade.

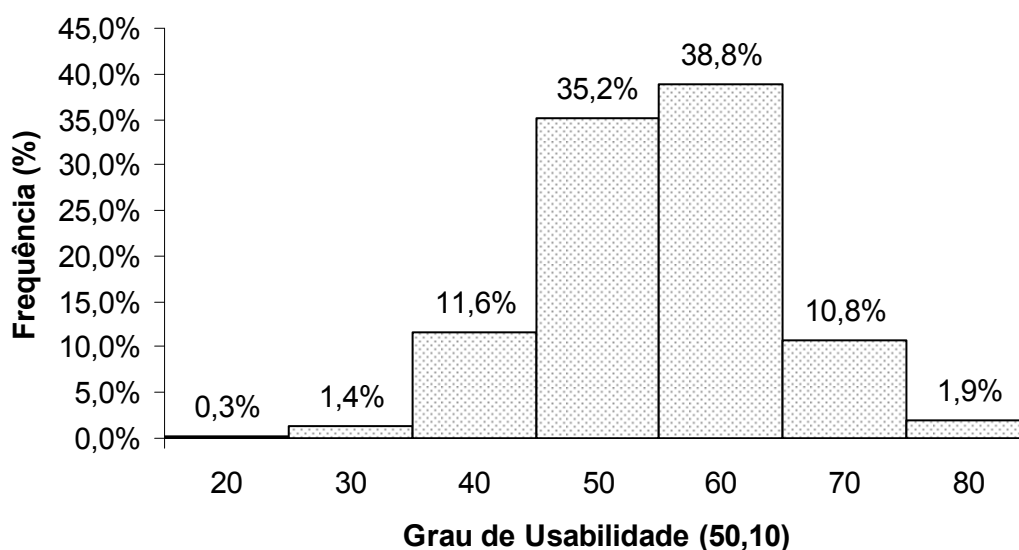


Figura 4.6 - Frequência de *sites* em cada nível da escala do grau de usabilidade na métrica (50,10)

O histograma da Figura 4.6 mostra que a maioria dos *sites* analisados possui grau de usabilidade em torno de 60, cerca de 51,5%, e os níveis extremos 20 e 80 representam respectivamente 0,3% e 1,9%.

4.5 Considerações Finais do Capítulo

Este capítulo teve como objetivo analisar a consistência da ferramenta utilizada (TRI) com base em seus resultados e no contexto prático e teórico de usabilidade. E a partir daí construir uma escala de usabilidade baseada nos itens elaborados.

Quando o resultado de uma modelagem recebe respaldo prático positivo existe grande chance de se estar no caminho certo, no que diz respeito a metodologia adotada bem com a ferramenta empregada. É exatamente esta situação verificada neste capítulo.

Nas três primeiras seções deste capítulo foram apresentados e discutidos os resultados estatístico obtidos com o software BILOG MG[®]. Dentre estes, os parâmetro de estatística descritiva (número de respondentes, número de acerto, correlação de Pearson e correlação bisserial) que são utilizados como valores iniciais para estimação dos parâmetros dos itens (parâmetro de discriminação “a”, parâmetro de dificuldade “b”) e estimação da proficiência

dos *sites* analisados. Nesta primeira etapa foram considerados todos os 44 itens elaborados. A partir desta análise foi possível identificar estatisticamente alguns itens com problemas de consistência com o conjunto, ou seja, não parecem estar medindo o mesmo traço latente que os demais, neste caso o grau de usabilidade. Estes itens com problema, foram eliminados do construto. A eliminação foi feita com base no parâmetro de discriminação, sendo que itens com “a” inferior a 0,700 foram desconsiderados do conjunto. Uma análise destes itens eliminados sob o ponto de vista de usabilidade mostrou que de fatos existiam problemas de ordem teórica e/ou prática com estes itens, confirmando a validade do instrumento. Doze dos 44 itens foram eliminados, restando, portanto, 32 itens para a construção definitiva da escala. Foi realizada a re-estimação dos parâmetros, agora com os 32 itens restantes. Os valores estimados dos novos parâmetros sofreram pouca alteração em relação aos do conjunto inicial, o que comprova a robustez do construto.

A função informação total mostrou que o construto é ideal para medir *sites* com baixa usabilidade

A construção da escala foi feita através da determinação de itens âncora e níveis âncora na métrica (50,10). Foram identificados 4 níveis âncora e 20 itens âncora ou “quase âncora”. O Quadro 4.11 traz uma síntese dos resultados obtidos, apresentando a descrição dos 32 itens juntamente com seus parâmetros e os níveis âncora na escala (50,10).

Item		Parâmetro		Grau de Usabilidade (Θ) (50,10)										
		a	b	0	10	20	30	40	50	60	70	80	90	100
02	Homepage deixa claro o que o site faz, sem precisar usar a rolagem?	0,076	34,59	0,07	0,13	0,25	0,41	0,60	0,76	0,87	0,94	0,97	0,99	0,99
04	As imagens, botões ou palavras clicáveis apresentam uma forma diferenciada quando são selecionadas?	0,089	35,37	0,04	0,10	0,20	0,38	0,60	0,78	0,90	0,96	0,98	0,99	1,00
06	As opções principais do site estão visíveis?	0,082	20,61	0,15	0,29	0,49	0,68	0,83	0,92	0,96	0,98	0,99	1,00	1,00
10	A disposição dos objetos de interação de uma caixa de diálogo segue uma ordem lógica?	0,079	40,98	0,04	0,08	0,16	0,30	0,48	0,67	0,82	0,91	0,96	0,98	0,99
11	Os rótulos de campos começam com uma letra maiúscula, e as letras restantes são minúsculas?	0,080	34,35	0,06	0,13	0,24	0,41	0,61	0,78	0,89	0,95	0,97	0,99	0,99
12	O site possui opção de acesso com outras línguas?	0,078	96,85	0,00	0,00	0,00	0,01	0,01	0,03	0,05	0,11	0,21	0,37	0,56
14	Títulos estão alinhados a esquerda?	0,104	27,02	0,06	0,14	0,32	0,58	0,79	0,92	0,97	0,99	1,00	1,00	1,00
15	Parágrafos de texto são separados?	0,093	20,80	0,13	0,27	0,48	0,70	0,86	0,94	0,97	0,99	1,00	1,00	1,00
16	As palavras aparentemente clicáveis são de fato clicáveis ?	0,097	51,77	0,01	0,02	0,04	0,11	0,24	0,46	0,69	0,85	0,94	0,98	0,99
17	Os títulos de telas, janelas e caixas de diálogo estão no alto, centrados ou justificados à esquerda?	0,142	25,00	0,03	0,11	0,33	0,67	0,89	0,97	0,99	1,00	1,00	1,00	1,00

(Continua ...)

(Continuação ...)

Item		Parâmetro		Grau de Usabilidade (Θ) (50,10)										
		a	b	0	10	20	30	40	50	60	70	80	90	100
19	Os resultados de busca permitem classificação por outros critérios além de custo?	0,129	54,52	0,00	0,00	0,01	0,04	0,13	0,36	0,67	0,88	0,96	0,99	1,00
20	Listas longas apresentam indicadores de continuação, de quantidade de itens e de páginas?	0,075	47,18	0,03	0,06	0,12	0,22	0,37	0,55	0,72	0,85	0,92	0,96	0,98
21	O preço de um produto consta ao lado da imagem ou do link do produto?	0,118	24,56	0,05	0,15	0,36	0,66	0,86	0,95	0,98	1,00	1,00	1,00	1,00
22	Existe uma orientação ao usuário quanto ao restante do site?	0,091	51,85	0,01	0,02	0,05	0,12	0,25	0,46	0,68	0,84	0,93	0,97	0,99
23	A maioria dos produtos possui informações sobre eles?	0,115	30,54	0,03	0,09	0,23	0,48	0,75	0,90	0,97	0,99	1,00	1,00	1,00
25	É possível ampliar as fotos dos produtos para visualizar detalhes?	0,098	40,51	0,02	0,05	0,12	0,26	0,49	0,72	0,87	0,95	0,98	0,99	1,00
26	Em produtos em que existam mais de uma perspectiva, é possível visualizar todas as perspectivas?	0,109	66,82	0,00	0,00	0,01	0,02	0,05	0,14	0,32	0,59	0,81	0,93	0,97
27	Os grupos de botões de comando estão dispostos em coluna e à direita, ou em linha e abaixo dos objetos aos quais estão associados?	0,094	21,55	0,12	0,25	0,46	0,69	0,85	0,93	0,97	0,99	1,00	1,00	1,00
28	Quando há rolagem, não existe elementos de design que pareçam com marcadores de final de página?	0,111	21,34	0,09	0,22	0,46	0,72	0,89	0,96	0,99	1,00	1,00	1,00	1,00
29	Todos os campos e mostradores de dados possuem rótulos identificativos?	0,105	28,51	0,05	0,12	0,29	0,54	0,77	0,91	0,96	0,99	1,00	1,00	1,00
30	O botão de finalização de compra está no final da lista?	0,075	5,10	0,41	0,59	0,75	0,87	0,93	0,97	0,98	0,99	1,00	1,00	1,00
31	É possível saber os custos totais antes de fazer cadastro?	0,087	29,51	0,07	0,15	0,30	0,50	0,71	0,86	0,93	0,97	0,99	0,99	1,00
34	No preenchimento de um formulário, é informado a forma de preenchimento?	0,109	44,65	0,01	0,02	0,06	0,17	0,38	0,64	0,84	0,94	0,98	0,99	1,00
35	Os dados obrigatórios são diferenciados dos dados opcionais de forma visualmente clara?	0,076	38,55	0,05	0,10	0,20	0,34	0,53	0,70	0,84	0,92	0,96	0,98	0,99
38	As mensagens de erro estão isentas de abreviaturas e/ ou códigos gerados pelo sistema operacional?	0,104	20,41	0,11	0,25	0,49	0,73	0,88	0,96	0,98	0,99	1,00	1,00	1,00
39	Todas as páginas possuem os mesmos layouts e exibem ao usuário as mesmas características?	0,128	33,63	0,01	0,05	0,15	0,39	0,69	0,89	0,97	0,99	1,00	1,00	1,00
40	O logotipo da empresa está no canto superior esquerdo em todas as páginas do site?	0,154	32,54	0,01	0,03	0,13	0,40	0,76	0,94	0,99	1,00	1,00	1,00	1,00
41	Existe um link de um único clique que conduz a homepage?	0,133	29,85	0,02	0,07	0,21	0,50	0,79	0,94	0,98	1,00	1,00	1,00	1,00
42	Qualquer ação do usuário pode ser revertida através da opção DESFAZER?	0,144	37,95	0,00	0,02	0,07	0,24	0,57	0,85	0,96	0,99	1,00	1,00	1,00
43	O site permite navegação em suas páginas em apenas uma janela, ou seja, não há abertura de novas janelas em meio a navegação?	0,111	33,06	0,03	0,07	0,19	0,42	0,68	0,87	0,95	0,98	0,99	1,00	1,00
44	Os links já visitados mudam de cor?	0,069	94,20	0,00	0,00	0,01	0,01	0,02	0,05	0,09	0,16	0,27	0,43	0,60

Quadro 4.11 – Itens do construto final e seus respectivos parâmetros e níveis na escala de usabilidade (50,10)

5 CONCLUSÕES E RECOMENDAÇÕES

5.1 Conclusões

O construto proposto visa somar objetividade as práticas de avaliação de usabilidade e contribuir para a difusão da Teoria de Resposta ao Item no contexto comercial. A proposta não elimina nenhuma metodologia de avaliação de usabilidade em *sites* comerciais, apenas acrescenta uma visão objetiva a estas.

O objetivo principal do presente trabalho foi elaborar um construto para medir usabilidade em *sites* de *e-commerce* tendo como base requisitos de usabilidade e fazendo uso da Teoria da Resposta ao Item. Isto foi obtido através da identificação dos seguintes aspectos de usabilidade: Design da página, Navegação, Busca e informações relacionadas, e Suporte a tarefa de compra. Dentro destes aspectos foram extraídas questões objetivas com o intuito de avaliar a eficiência e produtividade em uma interação usuário-interface, formando assim um questionário com 44 itens. Este questionário foi aplicado em 361 *sites*. Sua validação se deu através da convergência entre os resultados obtidos com a Teoria da Resposta ao Item e os respaldos teóricos e práticos no contexto de usabilidade em *sites* comerciais. A análise eliminou alguns itens por falta de consistência ao conjunto, resultando assim em um construto com 32 itens.

A principal contribuição do estudo está na criação de uma escala padronizada do grau de usabilidade, que permite a comparabilidade entre *sites* de *e-commerce* e também um acompanhamento longitudinal dos avanços de determinado *site* em termos de usabilidade, baseado no índice de respostas destes. Os resultados obtidos podem ser utilizados pela empresa de *e-commerce*, de uma forma prática, para eleger prioridades no desenvolvimento ou reformulação de seu *site* auxiliando assim na tomada de decisão e/ou melhoria do *site*.

Além disso, a criação da escala padronizada identificou quatro níveis âncora, sendo possível verificar as características de usabilidade predominantes. Permitindo assim, verificar quais características agrega mais usabilidade e quais são as mais básicas. As características mais básicas, segundo a análise, estão relacionadas ao design da página, mais especificamente os aspectos que tratam de padrões visuais, legibilidade, modelos visuais, etc. Estes aspectos, no contexto de usabilidade, podem dificultar a interação. Entretanto, dificilmente impedirão o

usuário de realizar sua tarefa. As características voltadas para a navegabilidade e elementos de busca mostraram-se aspectos de nível mais alto na escala de grau de usabilidade, principalmente por acrescentarem aos aspectos visuais o poder de interatividade que pode significar a impossibilidade de realização da tarefa.

Sendo assim, o construto foi validado no contexto de usabilidade, tanto no ponto de vista estatístico quanto do teórico-prático. Contribuindo para um melhor entendimento das relações de interação entre consumidor e *sites* comerciais e difundindo estatísticas robustas de avaliação de desempenho como a Teoria da Resposta ao Item.

Do ponto de vista particularizado de usabilidade, os resultados finais demonstraram a potencialidade da Teoria da Resposta ao Item para avaliação desta, podendo assim representar uma nova via de avaliação, com a vantagem de eliminar inconvenientes inerentes as técnicas usuais (subjetividade e alto custo).

5.2 Recomendações para Trabalhos Futuros

O desenvolvimento e os resultados deste trabalho abrem várias outras oportunidades de pesquisa no campo de usabilidade, comércio eletrônico e Teoria da Resposta ao item (TRI), dentre estas se podem citar:

- Ampliação do conjunto de itens, a fim de abranger a escala desenvolvida no sentido de maior grau de usabilidade. Esta ampliação pode ser feita explorando conceitos mais particulares de usabilidade de interfaces. Para isso, um novo processo de equalização utilizando os itens já existente e os novos itens deverá ser realizada.
- O presente trabalho fez uso de um dos modelos matemáticos da TRI, o modelo logístico de dois parâmetros. Sendo assim, recomenda-se a exploração de outros modelos neste contexto. Verificou-se que dentro do conceito comercial, os aspectos relacionados a publicidade on-line, não necessariamente seguem um comportamento acumulativo (característica básica para aplicação do modelo logístico de dois parâmetros), pois envolvem outros fatores, como maturidade tecnológica, inovação, interesses diversos, necessidade de exposição entre outras. Talvez o modelo de desdobramento possa modelar de forma mais real esta característica.
- Neste trabalho são tratados apenas aspectos objetivos, sem influência direta do usuário. Desta forma, recomenda-se a ampliação desta abordagem a aspectos subjetivos, relacionados

a satisfação do usuário e aspectos ergonômicos de uso. Para isso é necessário estudar a interação de grupos de usuários e encontrar no conjunto de modelos da TRI o que melhor modela esta avaliação usuário-tarefa-interface.

- Dentro de mercado de comércio eletrônico, como foi descrito no capítulo 2, existem sub modalidades, como é o caso do comércio eletrônico móvel (*m-commerce*) que faz uso de dispositivos móveis, e vem ganhando cada vez mais espaço no mercado digital. Estes dispositivos apresentam características particulares de uso e interação representando assim uma possibilidade de desenvolvimento de um construto para medição de usabilidade utilizando a TRI.

- O processo de compra em um site comercial envolve, além da interação com o site, outros fatores, como logística, pós-venda, custos entre outros. Desta forma, o desenvolvimento de um construto capaz de avaliar e medir estes fatores, e colocá-los em uma escala que permita comparabilidade, torna-se um grande desafio.

REFERÊNCIAS

- ALBERTIN, A. L. **Comércio eletrônico : modelo, aspectos e contribuições de sua aplicação** / Alberto Luiz Albertin. Colaboração de Rosa Maria de Moura. 5. Ed. São Paulo : Atlas, 2004.
- ALEXANDRE, J. W. C.; ANDRADE, D.F.; VASCONCELOS, AP.; ARAUJO, A M S. Uma proposta de análise de um construto para a medição dos fatores críticos da gestão pela qualidade através da teoria da resposta ao item. **Gestão & Produção**. v.9, n.2, p.129-141.
- AMENTO, B.; HILL, W.; TERVEEN, L.; HIX, D.; JU, P. **An empirical evaluation of user interfaces for topic management of web sites**. Proceedings of CHI'99, p. 552-559, 1999.
- ANDERSEN, E.B. A goodness of fit test for the Rasch model. **Psychometrika**, v.38, p.123-140. 1973.
- 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.
- ANDRICH, D. A rating formulation for ordered response categories. **Psychometrika**, v.43, p. 561-573. 1978.
- APPLE COMPUTER, Inc. **Macintosh Human Interface Guidelines** – Addison-Weslwy Company, 1992.
- ARAGÃO, C. R. V. **A Percepção do usuário sobre o fator usabilidade das páginas da web voltadas para o comércio eletrônico**, Dissertação de mestrado - Universidade Federal de Santa Catarina, Centro tecnológico. Programa de Pós-Graduação em Engenharia de Produção, Florianópolis, SC, 2001.
- BAKER, F. B. **The Basics of Item Response Theory**. 2 ed. USA: ERIC Clearinghouse on Assessment and Evaluation, 2001. Disponível em: <http://edres.org/irt/> Acessado em 12 agosto 2008
- BAKER, F. B., KIM, S. **Item Responde Theory: parameter estimation techniques**. Second edition, revised and expanded. New York: Marcel Dekker, 2004.
- BASTIEN, J. M. C., SCAPIN, D. L. **Ergonomic criteria for the evaluation of humam-computer interfaces**. Rapport technique de l'INRIA. 1993. Disponível em : <http://www.inria.fr/rrrt/rt-0156.html>.
- BAYLEY, S. Measuring customer satisfaction". **Evaluation Journal of Australasia**, v. 1, n. 1, 2001.
- BAILEY, R.W.; KOYANI, S.J. **Searching vs. linking on the web: A summary of the research**. Health and Human Services Technical Report, 2004.
- BERNARD, M.; LARSEN, L. **What is the best layout for multiple-column Web pages?** Usability News, 3.2. 2001. Disponível em: <http://psychology.wichita.edu/surl/usabilitynews/3S/layout.htm>. Acessado em 01 set. 2008.
- BERNARD, M., LIAO, C.H., MILLS, M. **The effects of font type and size on the legibility and reading time of online text by older adults**. Proceedings of CHI 2002, 175-176, 2001. Disponível em <http://psychology.wichita.edu/hci/projects/elderly.pdf>. Acessado em 05 nov 2008.

- BIRBAUM, A. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In F. M. Lord and M. R. Novick. *Statistical Theories of Mental Test Scores*, MA: Addison-Wesley. goodness of fit test for the Rasch model. **Psychometrika**, v.38, p. 123-140, 1968.
- BITITCI, U. S. Modelling of performance measurement systems in manufacturing enterprises. **International Journal of Production Economics**, v.42, p.137-147, 1995.
- BLOCH, M.; PIGNEUR, Y.; SEGEV, A. **On the road of electronic commerce: a business value framework, gaining competitive advantage and some research issue**. Lausanne: Institut D'Informatique et Organisation, Ecole des Hautes Etudes Commerciales, Université de Lausanne, 1996.
- BOCK, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. **Psychometrika**, v.37, p. 29-51, 1972.
- BOCK, R. D.; AITKIN, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. **Psychometrika**, v.46, p. 443-459, 1981.
- BOCK, R.D.; LIEBERMAN, D. Fitting a response model for n dichotomously scored items. **Psychometrika**, v.35, p.179-197, 1970.
- BOCK, R. D.; ZIMOWSKI, M.F. **Multiple Group IRT**. In *Handbook of Modern Item Response Theory*. W. J. van der Linder and R.K. Hambleton Eds. New York: Springer-Verlag, 1997.
- BORGES, J. A., MORALES, I.; RODRÍGUEZ, N. J., **Guidelines for Designing Usable World Wide Web Pages**, CHI 96 Companion, Vancouver, BC Canada, p. 277-278, 1996.
- BYRNE, M.D.; ANDERSON, J.R.; DOUGLASS, S.; MATESSA, M. **Eye tracking the visual search of click-down menus**. Proceedings of CHI'99, p.402-409, 1999.
- CAMERON, D. **Electronic commerce: the new business platform of the Internet**. Charleston: Computer Technology Research Corp, 1997.
- CAMPBELL, C.S.; MAGLIO, P.P. Facilitating navigation in information spaces: Road signs on the World Wide Web. **International Journal of Human-Computer Studies**, v.50, p.309-327, 1999.
- CELLA, D. CHANG, C.H., A discussion of Item Response Theory and it's application in health status assessment. **Medical Care**, v.38 p.66-72, 2000.
- CHAPARRO, B.S.; BERNARD, M.L. **Finding information on the Web: Does the amount of whitespace really matter?** Proceedings of the Tenth Annual Usability Professionals' Association Conference, 2001.
- CHIN, J. P., DIEHL, V. A., NORMAN, K. L. **Development of an instrument measuring user satisfaction of the human-computer interface**. *Proc. ACM CHI88 Con!* (Washington, DC, 15-19 May), 213-218. 1988.
- COSTA, M. B. F. **Técnica derivada da Teoria de Resposta ao Item (TRI) aplicada ao setor de serviços**. Dissertação de Mestrado, UFPR, Curitiba, 2001.
- CULLEN, A. J., WEBSTER, M.. A model of B2B e-commerce, based on connectivity and purpose. **International Journal of Operation & Production Management**, v. 27 n. 2, p. 205-225, 2007.
- CURI, M. **Análise de questionários com itens constrangedores**. Tese de Doutorado. IME/USP. São Paulo, 2006.

CYBIS, W. A. **Abordagem Ergonômica para IHC**. Apostila LabIUtil, Universidade Federal de Santa Catarina, 1997.

CYBIS, W. **Engenharia ergonômica de usabilidade de interfaces humanocomputador**. Apostila para o Curso de Pós-Graduação em Engenharia de Produção – Universidade Federal de Santa Catarina, 2002.

CYBIS, W. **Ergonomia e Usabilidade** : conhecimentos, métodos e aplicações / Walter Cybis, Adriana Holtz Betiol, Richard Faust. São Paulo : Novatec Editora, 2007.

DAS, J.; HAMMER, J. Which doctor? Combining vignettes and item response to measure clinical competence. **Journal of Development Economics** v.78, p.348-383, 2005.

DAVIS, F., .Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology,. **MIS Quarterly**, v. 13, n.3 p.391-340, 1989.

DELFMANN, W.; ALBERS, S.; GEHRING, M., The impact of electronic commerce on logistics service providers. **International Journal of Physical Distribution & Logistics Management**, v.32 n.3, p.203-222.

DeROOS, Y.; MEARES, P. A. “Aplication of Rasch analysis: exploring differences in depression between african - american and white children”. **Journal of Social Service Research**, v. 23, n 3/4, p. 93-107, 1998.

DETWEILER, M.C., OMANSON, R.C. **Ameritech Web Page User Interface Standards and Design Guidelines**. SBC, 1996.

DIAS, C. A.. Grupo Focal: técnica de coleta de dados em pesquisas qualitativas. **Informação e Sociedade**, v.10, n.2, 2000.

DIX, A.J.; FINLAY, J.E.; ABWOD, G.D.;BEALE, R. **Human-Computer Interaction**. Prentice-Hall, NJ, 1998.

DUMAIS, S.; CUTRELL, E.; CHEN, H. **Optimizing search by showing results in context**. Proceedings of CHI 2001, p. 277-283, 2001.

DUMAS, J.S.; REDISH, J.C., **A Practical Guide to Usability Testing**. Ablex, Norwood, NJ. 1994.

E-BIT. **Web Shoppers**. 18^a ed. 2008. Disponível em: <http://www.webshoppers.com.br/webshoppers.asp> Acessado em 19 jan. 2009.

E-BIT. **Web Shoppers**. 19^a ed. 2009. Disponível em: <http://www.webshoppers.com.br/webshoppers.asp> Acessado em 19 mar. 2009.

EDELEN, M. O.; REEVE, B. B. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. **Qual Life Res** v.16 p.05–18. 2007

EGGER, F.N. **Human factors in Electronic Commerce**: Making systems appealing, usable and trustworthy. In Proceedings of the Graduate Students consortium & educational symposium, 12th Bled International E-Commerce Conference, Bled, Slovenia, June 1999.

ELLIS, R.D.; KURNIAWAN, S.H. Increasing the usability of online information for older users: A case study of participatory design. **International Journal of Human-Computer Interaction**, v.12 n.2, p. 263-276, 2000.

EMBRETSON, S.; REISE, S. P. **Item Response Theory for Psychologists**. New Jersey: Lawrence Erlbaum Associates, Inc. Publishers, 2000.

FERGUSON, G. A. Item selection by the constant process, **Psychometrika**, v.7 n.1, 1942.

- FISCHER, G. H. The linear logistic test model as an instrument in educational research. *Acta Psychologica*. v. 37 n.6 p. 359-374. 1973.
- FREUND, J.E.; SIMON, G.A. **Estatística básica**: economia, administração e contabilidade. 9 ed. Porto Alegre: Bookman, 2000.
- GABBOTT, M.; HOGG, G. **Consumers and Services**, John Wiley & Sons, Chicester, UK, 1998.
- GIL, A. C. **Como elaborar projetos de pesquisa**. 4º ed São Paulo: Atlas, 2002.
- GÓIS, S.M. Algumas considerações sobre comércio eletrônico. **Conjuntura & Planejamento**, Salvador: SEI/CPE, n.71, p. 25-26, abr., 2000.
- GRANGER, C V.; DEUTSCH, A.. Rasch analysis of the functional independence measure (FIMTM) mastery test. **Arch Phys med Rehabil**, v.79, p.52-57, 1998.
- GUEWEHR, K. **Teoria da resposta ao item na avaliação de qualidade de vida de idosos**. Dissertação de mestrado - Universidade Federal do Rio Grande do Sul, Faculdade de Medicina, Programa de Pós-Graduação em Epidemiologia, RS, 2007.
- GÜNTHER, H. **Como elaborar um questionário** (Série: Planejamento de Pesquisa nas Ciências Sociais, nº 01). Brasília, DF: UnB, Laboratório de Psicologia Ambiental, 2003.
- 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.
- HAMBLETON, R.K., SWAMINATHAN, H., **Item response theory**: Principles and applications. Norwell, MA: Kluwer Academic Publishers, 1985.
- HAMBLETON, R.K., SWAMINATHAN, H., ROGERS,H.J. **Fundamentals of item response theory**. Newbury Park, CA: Sage, 1991.
- HEDGES, L. V.; VEVEA, J. L. **A study of equating in NAEP. Paper presented at The NAEP Validity Studies Panel**. Palo Alto : American Institutes for Research. 1997.
- HFDG – **Human Factors Design Guide Update** (Report Number DOT/FAA/CT-96/01): A Revision to Chapter 8 – Computer Human Interface Guidelines. US Federal Aviation Administration Technical Center – DOT/FAA/ACT-01/08, Washington, DC, 2001.
- HHS - U.S. Department of Health and Human Services . **The Research-Based Web Design & Usability Guidelines** (Guidelines). 2006 disponível em: <http://www.usability.gov/guidelines/index.html>. Acesso em 20 de abril 2008.
- HIX, D.; HARTSON, H. R. **Developing User Interfaces**: Ensuring Usability Through Product and Process. New York: John Wiley and Sons, 1993.
- ISO 9241. **Ergonomic of human-system interaction**. Part 110 Dialogue principles; Draft International Standard ISO, 2006.
- ISO 9241. **Ergonomic requirements for office work with visual display terminals (VDTs)**. Part 3 Visual display requirements; Draft International Standard ISO, 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.

- JEFFRIES, R., MILLER, J.R., WHARTON, C., UYEDA, K.M. **User interface evaluation in the real world A comparison, of four techniques**. Proc. ACM CHI'91 (New Orleans, LA, 27 April–2 May), p. 119–124. 1991.
- JORDAN, P. **An introduction to usability**. Taylor & Francis, London. 1998.
- KALAKOTA, R., WHINSTON, A. B. **Eletronic Commerce: a manager's guide**. Addison Wesley INC, 1997.
- KIERAS, D. E.; POLSON, P. G. An approach to the formal analysis of user complexity. **International Journal of Human-Computer Studies** v.51 n.2 p.405-434, 1999.
- KIM, S. STOEL, L. Dimensional hierarchy of retail website quality, **Information & Management**, v. 41 n. 5, p. 619-633, 2004.
- LAWLEY, D. N. **On problems connected with item selection and test construction**. Proceedings of the Royal Society of Edinburgh 61, p. 273-287, 1943.
- LAZARSELD, P.F. Chapter 10 and 11 in S.A. Stouffer et al. (eds.). **Studies in Social Psychology in World War II: v. 4. Measurement and Prediction**. Princeton, NJ: Princeton University Press. 1950.
- LEBAS, M. J. Performance measurement and performance management. **International Journal of Production Economics**, v.41, p.23-35, 1995.
- 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.
- LINACRE, J. M. Sample size and item calibration stability, **Rasch Measurement Transactions**, v.7 n.4, p.328, 1994.
- LOHMAN, C., FORTUIN, L.; WOUTERS, M. Designing a performance measurement system: a case study, **European Journal of Operational Research**, v. 156 n. 2, p. 267-286, 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.
- LORD, F. M. Estimation of latent ability and item parameters when there are omitted responses. **Psychometrika**, v.39, p247-264. 1974.
- LORD, F. M.; NOVICK, M. R. **Statistical Theories of Mental Test Score**. Reading: Addison-Wesley, 1968.
- LYNCH, P.J.; HORTON, S. **Web Style Guide 2º ed**. New Haven, CO: Yale University Press. 2002 disponível em <http://www.webstyleguide.com/index.html?/contents.html>. acessado em 05 de julho 2008.
- MADER, M. J. . Avaliação Neuropsicológica - Aspectos históricos e Situação Atual. **Psicologia Ciência e Profissão**, v. 16, n. 3, 1996
- MASTERS, G. N. A Rasch model for partial credit scoring. **Psychometrika**, v. 47, p.149-174. 1982.
- MATIAS, M., **CheckList** : Uma Ferramenta de Suporte à Avaliação Ergonômica de Interfaces, Dissertação de Mestrado, Universidade Federal de Santa Catarina, Centro

tecnológico. Programa de Pós-Graduação em Engenharia de Produção, Florianópolis, SC, 1995.

MAYHEW, D.J. **The Usability Engineering Lifecycle**: a practitioner's handbook for user interface design. Morgan Kaufmann Publishers, CA, 1999.

METZ, S.M.; WYRWICH, K.W.; BABU, A.N.; KROENKE, K.; TIERNEY, W.M.; WOLINSKY, F.D., A comparison of traditional and Rasch CUT points for assessing clinically important change in health-related quality of life among patients with asthma. **Quality of Life Research**, v. 15, p.1639-1649, 2006.

MILONE, G. **Estatística Geral e Aplicada**. São Paulo: Pioneira Thomson Learning, 2004.

MINOCHA, S.; PETRE, M.; TZANIDOU, E.; VAN DIJK, G.; MILLARD, N; ROBERTS, D.; DAY, B. GASSMAN, N.; TRAVIS, D. **Evaluating E-Commerce Environments**: Approaches to Cross-disciplinary Investigation. CHI 2006, April p. 22–27, Montréal, Québec, Canadá, 2006.

MISLEVY, R.J. Bayes modal estimation in item response models. **Psychometrika**, v.51, p.177–195. 1986.

MISLEVY, D. J., BOCK, R. D. BILOG: **Item analysis and test scoring with binary logistic models** [Computer program]. Chicago: Scientific Software, 1990.

MONK, A.; WRIGHT, P.; HABER, J.; DAVENPORT, L.. **Improving your human-computer interface**: a practical technique. New York, Prentice Hall, 1993.

MONTGOMERY, D.C.;RUNGER,G.C. **Estatística aplicada e probabilidade para engenheiros**, 2º edição, LTC, Rio de Janeiro, RJ, 2003

MORANDINI, M. **Ergo-Monitor: Monitoramento da Usabilidade em Ambiente Web por meio da Análise de Arquivos de Log** Tese Doutorado - Universidade Federal de Santa Catarina, Centro tecnológico. Programa de Pós-Graduação em Engenharia de Produção, Florianópolis, SC, 2007.

MOSIER, C. I. A modification of the method of successive intervals. **Psychometrika**, v. 7 n. 1 p. 19-29, 1942.

MOSIER, C. I. A psychometric study of meaning. **Journal of Social Psychology**, v.13, p.123-140. 1941.

MOTTA, P. C. **Serviços: pesquisando a satisfação do consumidor**. Rio de Janeiro: Imprinta Express, 1999.

MURAKI, E. A generalized partial credit model: Application of an EM algorithm. **Applied Psychological Measurement**, v.16, p.159-176. 1992.

MURAKI, E.; BOCK, R.D. PARSCALE: **IRT based test scoring and item analysis for graded open-ended exercises and performance tasks**. Chicago: Scientific Software Int. 1993.

MURAKI, E.; ENGELHARD, G. Full information item factor analysis: Application of EAP score. **Applied Psychological Measurement**, v.9, p417-430, 1985.

NIELSEN, J. **Accessible design for users with disabilities**. October, 1996. Disponível em: <http://www.useit.com/alertbox/9610.html>. Acessado em 20 de Nov. de 2008.

NIELSEN, J. **Projetando Websites**. Rio de Janeiro: Elsevier , 2000.

- NILSEN, J. **Search: Visible and simple**. May 2001. Disponível em: <http://www.useit.com/alertbox/20010513.html>. Acessado em 20 de Nov. de 2008.
- NIELSEN, J. **The need for speed**. March, 1997. Disponível em <http://www.useit.com/alertbox/9703a.html> Acessado em 20 de Nov. de 2008.
- NIELSEN, J. **Usability Engineering**. California : Morgan Kaufmann , 1993.
- NIELSEN, J.; LORANGER, H. **Prioritizing Web Usability**. California : New Riders, California, 2006.
- NIELSEN, J.; MACK, Robert. **Usability Inspection Methods**. New York: John Wiley & Sons, 1994.
- NIELSEN, J.; MOLICH, R.; SNYDER, C.; FARRELL, S. **E-Commerce user experience: High-level strategy**. Nielsen Norman Group, 2001
- NIELSEN, J.; TAHIR, M. **Homepage Usability: 50 Sites Deconstructed**. Indianapolis, IN: New Riders Publishing, 2002.
- NORMAN, D. A. Design rules based on analyses of human error. **Communications of the ACM** v.26, n.4, p.254-258. 1983.
- NOVAES, A. G. **Logística e gerenciamento da cadeia de distribuição: estratégia, operação e avaliação**. 2ed. – Rio de Janeiro: Elsevier, 2004.
- NYGREN, E.; ALLARD, A. **Between the clicks: Skilled users scanning of pages**. Proceedings of the 2nd Conference on Human Factors and the Web, 1996. Disponível em <http://www.microsoft.com/usability/webconf/nygren.rtf>. Acessado em 20 nov. 2008.
- O'CONNOR, T. **Lecture notes for research methods**, 2007. Disponível em <http://www.apsu.edu/oconnort/3760/3760lects.htm> Acessado em 10 dez. 2008.
- ORLANDO, M.; MARSHALL, G. N. Differential item functioning in a Spanish translation of the PTSD checklist: Detection and evaluation of impact. **Psychological Assessment**, v.14 n.1, p.50–59. 2002.
- PASQUALI, L. Princípios de elaboração de escalas psicológicas. **Revista Psiquiatria de Clínica**, v. 25, n. 5, p. 206-21, 1998.
- PASQUALI, L. **Psicometria: teoria dos testes na psicologia e na educação**. Petrópolis, RJ: Vozes, 2003.
- PASQUALI, L. **Psicometria: teoria e aplicações**. Brasília : Editora Universidade de Brasília, 1997.
- PASQUALI, L.; PRIMI, R. Fundamentos da Teoria da Resposta ao Item – TRI. **Avaliação Psicológica**, v.2 n.2, p. 99-110, 2003.
- PETRE, M.; MINOCHA, S.; ROBERTS, D. Usability beyond the website: An Empirically-Grounded Service Quality Toolbox for E-Commerce Environments. **Behaviour and Information Technology**, v.25, n.3, p 189-203, 2006.
- PIOLAT, A.; ROUSSEY, J.Y.; THUNIN, O. Effects of screen presentation on text reading and revising. **International Journal of Human Computer Studies**, v.47, p.565-589, 1998.
- POLLOCK, A.; HOCKLEY, A. **What's wrong with Internet searching**. Proceedings of the 2nd Conference on Human Factors and the Web, 1996. Disponível em: <http://www.microsoft.com/usability/webconf/pollock.rtf>. Acessado em 20 nov. 2008.

- RASCH, G. **Probabilistic Models for Some Intelligence and Attainment Tests**. Copenhagen, Denmark: Danish Institute for Educational Research. 1960.
- REASON, J. **Human Error**. Cambridge University Press, Cambridge, UK. 1990.
- RECKASE, M. D. Unifactor latent trait models applied to multifactor tests: Results and implications. **Journal of Educational Statistics**, v.4, p.207-230. 1979.
- RICHARDSON M.W. The relationship between difficulty and the differential validity of a test. **Psychometrika**. v.1 p.33-49. 1936.
- ROSENFELD, L.; MORVILLE, P. **Information Architecture for the World Wide Web** (second edition). Sebastopol, CA: O'Reilly, 2002.
- SAMEJINA, F. **Estimation of latent ability using a response pattern of graded scores**. Psychometrika Monograph Supplement, 1, 1969.
- SANTOS, A. R.. **Metodologia científica: a construção do conhecimento**. Rio de Janeiro: DP & A, 1999.
- SANTOS, R. L. G. dos, **Usabilidade de interfaces para sistemas de recuperação de informação na web: estudo de caso de bibliotecas on-line de universidades federais brasileiras** Tese Doutorado – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Artes e Design – Rio de Janeiro – RJ, 2006.
- SCHUMAN, H. KALTON, G. **Survey methods**. In G. Lindezey; E. Aronson (Eds.), Handbook of social psychology, 3^oed.. V. 1 New York: Random House, p 635-697, 1985.
- SCHMITT, N. Uses and abuses of coefficient alpha. **Psychological Assessment**, v.8, p.350-353, 1996.
- SENDERS, J. W., MORAY, N. P. **Human Error: Cause, Prediction, and Reduction**. Erlbaurn, Hillsdale, NJ. 1991.
- SHAW, C.; IVENS, J. **Building Great Customer Experiences**, Palgrave Macmillan, 2002.
- SHNEIDERMAN, B., **Designing the User Interface – Strategies for Effective Human-Computer Interaction**, Los Angeles: Addison-Wesley, 1987.
- SINGH, J. Tackling measurement problems with Item Response Theory: Principles., characteristics, and assessment, with an illustrative example. **Journal of Business Research**, v. 57, p. 184-208, 2004.
- SOARES, T.M. Utilização da teoria da resposta ao item na produção de indicadores sócio-econômicos. **Pesquisa Operacional**, Rio de Janeiro, v.25, n.1, 2005.
- SOMMER, B; SOMMER, R. **A practical guide to behavioral research: Tools and techniques** (4th ed). New York: Oxford U Press.
- SONG, J.H.. ZINKHAN, G.M. **Features of web site design, perceptions of web site quality, and patronage behaviour**, ACME 2003 Proceedings, pp. 106-14, 2003.
- SPENCER, S. G., **The strength of multidimensional item response theory in exploring construct space that is multidimensional and correlated**. A dissertation submitted to the faculty of Brigham Young University in partial fulfillment of the requirements for the degree of Doctor of Philosophy, November 19, 2004.
- SPYRIDAKIS, J.H. Guidelines for authoring comprehensible web pages and evaluating their success. **Technical Communication**, 47(3), 359-382, 2000.

- SPOOL, J.M.; SCANLON, T.; SCHROEDER, W.; SNYDER, C.; DeANGELO, T. **Web Site Usability: A Designer's Guide**. North Andover, MA: User Interface Engineering, 1997.
- STEINBERG, L., THISSEN, D.; WAINER, H. Validity. In H. Wainer (Ed), **Computerized Adaptive Testing: A Primer**. Mahwah, Jew Jersey: Lawrence Erlbaum. 2000.
- STEVENS, S.S. On the theory of scales of measurement. **Science**, v.103, p 221-263, 1946.
- STEVERSON, W. **Estatística aplicada à administração**. São Paulo: Harper & Row do Brasil, 1981.
- SWAMINATHAN, H.; GIFFORD, J.A. Bayesian Estimation in the Rasch model. **Journal of Educational Statistics**, v.7, p.175-192, 1982.
- SWAMINATHAN, H.; GIFFORD, J.A. Bayesian Estimation in the two-parameter logistic model. **Psychometrika**, v.50, p.349-364, 1985.
- SWAMINATHAN, H.; GIFFORD, J.A. Bayesian Estimation in the three-parameter logistic model. **Psychometrika**, v.51, p.589-601, 1986.
- TAVARES, H. R.; ANDRADE, D. F.; PEREIRA, C.A. Detection of determinant genes and diagnostic via item response theory. **Genetics and Molecular Biology**, v. 27, n. 4, p. 679-685, 2004.
- TERMAN, L. M. **The measurement of intelligence**. Boston: Houghton Mifflin. 1916.
- TEZZA, R. ; BORNIA, A. C. **Proposta de uma modelagem para mensuração de desempenho em e-commerce B2C**. In: V Simpósio de Excelência em Gestão e Tecnologia (SEGET), 2008, Resende, RJ. Anais do V Simpósio de Excelência em Gestão e Tecnologia (SEGET). Resende, RJ : Associação Educacional Dom Bosco, 2008. v. 1. p. 0-0.
- THISSEN, D. **MULTILOG: Multiple, categorical item analysis and test scoring using item response theory** [Computer program]. Chicago: Scientific Software. 1991.
- THISSEN, D., Steinberg, L., Gerrard, M. Beyond groupmean differences: The concept of item bias. **Psychological Bulletin**, v. 99 n.1, p118-128. 1986.
- THISSEN, D. **Estimation in multilog**. In M. du Toit (Ed.), IRT from SSI: Bilog-MG, multilog, parscale, testfact. Lincolnwood, IL: Scientific Software International. 2003.
- THURSTONE, L. L. A law of comparative judgments. **Psychological Review**, v.34, p.278-286, 1928.
- THURSTONE, L. L. Attitudes can be measured. **American Journal of Sociology**, v.33, p.529-554, 1927.
- TRUCKER, L.R. Maximum validity of a test with equivalent items. **Psychometrika** v.11, p.1-13. 1946.
- TSUTAKAWA, R.K.; JOHNSON, J.C. The effect of uncertainty of item parameter-estimation on ability estimates. **Psychometrika**, v.55, p.371-390. 1990.
- TUCKER, L.R. Maximum validity of a test with equivalent items. **Psychometrika**, v.11, p.1-13. 1946.
- USHIWATA, F.; VICENTE, R.S.; PRADO, A.A.; SILVA, M.B.; RIBEIRO, R.B., **As perspectivas do Comércio Eletrônico no Brasil e no mundo**. Janus, Lorena, ano 3, nº 4, 2006. Disponível em: <http://publicacoes.fatea.br/index.php/janus/article/viewFile/34/37>. acessado em 20 dez. 2008.

- VAN DER LINDEN, W. J.; HAMBLETON, R. K. **Handbook of Modern Item Response Theory**. New York: Springer- Verlag, 1997.
- VARGAS, V. do C. C. de. **Medida padronizada para avaliação de intangíveis organizacionais por meio da teoria da resposta ao item**. Tese Doutorado - Universidade Federal de Santa Catarina, Centro tecnológico. Programa de Pós-Graduação em Engenharia de Produção, Florianópolis, SC, 2007.
- VIDOTTO G.; BERTOLOTTI G.; CARONE M., F.; ARPINELLI F.; BELLIA V.; JONES P. W.; DONNER, C. F. A new questionnaire specifically designed for patients affected by chronic obstructive pulmonary disease: The Italian Health Status Questionnaire. **Respiratory Medicine**, v. 100, n. 5, p. 862-870, 2006.
- WANG, Y., TANG, T., TANG, J.E. An instrument for measuring customer satisfaction toward web sites that market digital products and services”, **Journal of Electronic Commerce Research**, v. 2 n. 3, p. 89-102, 2001.
- WILLIAMS, T.R. Guidelines for designing and evaluating the display of information on the Web. **Technical Communication**, v.47 n.3, p.383-396, 2000.
- WILSON, D.T., WOOD, R.; GIBBONS, R. TESTFACT: **Test scoring, item statistics, and item factor analysis**. Mooresville, IN: Scientific Software, 1991.
- WINCKLER, M. **Avaliação de usabilidade de sites Web**. In: IV Workshop sobre fatores humanos em sistemas de computação. Out. 2001, Florianópolis.
- WRIGHT, B. D.; DOUGLAS, G. A. Best procedures for sample-free item analysis. **Applied Psychological Measurement**, v.1, p.281-294, 1977.
- WU, I. Model Management system for IRT – based test construction decision support system. **Decision Support Systems**, v. 27, p. 443-458, 1999.
- ZIMOWSKI, M. F., MURAKI, E., MISLEVY, R. J., BOCK, R. D. **BILOG-MG**: Multiple-group IRT analysis and test maintenance for binary items. Chicago: Scientific Software International. 1996.

APÊNDICES

APÊNDICE A – Endereço dos sites avaliados

Endereço eletrônico dos Sites Avaliados		Endereço eletrônico dos Sites Avaliados	
1	http://www.angeloni.com.br	58	http://loja.trav.com.br/
2	http://www.big.com.br	59	http://www.somlivre.com/
3	http://www.calvi.com.br/	60	http://www.livrariasaraiva.com.br/
4	http://www.hippo.com.br	61	www.livrariaconvivencia.feesc.org.br/
5	http://www.paodeacucar.com.br	62	http://www.euroferragens.com.br/loja.phtml
6	http://www.marcos.com.br/index.php	63	http://www.altanatubes.com/index.asp
7	http://www.extra.com.br/	64	http://www.dutramaquinas.com.br/
8	http://www.bigsupermercados.com.br/index.php	65	http://sofrio.com.br/
9	http://www.esporteline.com.br/	66	http://www.ambientair.com.br/
10	http://www.netshoes.com.br/	67	http://www.poloar.com/
11	http://www.emporiumdosport.com.br/	68	http://www.kdpneus.com.br/
12	http://procorrer.com.br/	69	http://www.autoz.com.br/default.asp
13	http://www.penatilha.com.br/	70	http://www.funuta.com.br/index.php
14	http://www.corpoperfeito.com.br/	71	http://www.sbrío.com/
15	http://www.timesetorcidas.com.br/	72	http://loja.trav.com.br/
16	http://www.alfaperfumes.com.br/	73	http://www.batistinha.com.br/
17	http://www.levata.com.br/	74	http://www.adias.com.br/loja/
18	http://www.rumo.com.br/	75	http://www.pescawebshop.com.br/home.asp
19	http://www.sepha.com.br/	76	http://www.casaancora.com.br/
20	http://perfumariacuritiba.com.br/	77	http://www.rumo.com.br/
21	http://www.dotstore.com.br/	78	http://www.shoptime.com.br/
22	http://casasade.com.br/index.php	79	http://www.apetrexo.com.br/
23	http://www.donnaperfumes.com.br/	80	http://www.amercantil.com.br/
24	http://www.sacks.com.br/	81	http://www.magazinemail.com.br/Default.aspx
25	http://www.taco.com.br/	82	http://shop.megaomni.com/
26	http://www.camisariacolombo.com.br/	83	http://www.baccos.com.br/
27	http://www.rodeowest.com.br/	84	http://soparaalergicos.locaweb.com.br/index.php
28	http://shopping.banner-link.com.br/	85	http://www.audiovideoecia.com.br/
29	http://www.transpaper.com.br/index.php	86	http://www.mundomax.com.br/
30	http://timefox.com.br/loja/	87	http://www.jogmusic.com.br/index.php
31	http://www.mytime.com.br/	88	http://www.eletroni.com.br/
32	http://www.sportsonline.com.br/	89	http://www.moonshadows.com.br/loja/
33	http://www.wkshop.com.br/loja/default_random.asp	90	http://www.snetcommerce4.com.br/ecommerce_site/
34	http://www.vendamoda.com.br/vmvarejo.aspx	91	http://www.001shop.com.br/lojas/ultraeco.com.br/
35	http://www.joias10.com.br/v1/	92	http://www.perfumes.com.br/
36	http://www.sp165.com.br/	93	http://www.ingresso.com.br/
37	http://www.volletjoias.com.br/loja/produtos.aspx	94	http://www.efacil.com.br/
38	http://www.sandrarosamadalena.com.br/	95	http://www.comprafacil.com.br/
39	http://www.toymania.com.br/site/index.cfm	96	http://www.maniavirtual.com.br/
40	http://www.fofobrink.com.br/ecommerce_site/	97	http://www.casaavideo.com.br/
41	http://www.livrarianovaera.com.br/	98	http://www.assineabril.com/index.jsp
42	http://www.001shop.com.br/lojas/cheirodealecraim.com.br/	99	http://www.kalunga.com.br/
43	http://www.kalan.com.br/loja/	100	http://www.americanas.com/
44	http://www.lojamistica.com.br/	101	http://www.submarino.com.br/
45	http://www.ecotree.com.br/	102	http://www.pontofrio.com.br/
46	http://aromaemagia.com.br/	103	http://www.magazineluiza.com.br/
47	http://www.luamistica.com.br/	104	http://www.freeitalia.com.br/
48	http://www.maraviglia.com.br/	105	http://www.virtuose.com.br/Lojas/00000003/asp/FrameVirtuose.asp?IDLoja=3&I=257
49	http://www.goodgold.com.br/	106	https://www.rumo.com.br/
50	http://www.catmania.com.br/	107	http://www.007sexshop.com.br/
51	http://www.astrogeo.net/	108	http://www.asiashop.com.br/
52	http://www.netbolsas.com.br/	109	http://www.fastcommerce.com.br/sistema/home.asp?IDLoja=9513&1ST=1&Y=4210116747225
53	http://www.donnaacessorios.com.br/loja/	110	http://www.virtuose.com.br/Lojas/00000003/asp/FrameVirtuose.asp?IDLoja=3&I=6989
54	http://www.polimaia.com.br/easycommerce/home.asp	111	http://www.virtuose.com.br/Lojas/00000003/asp/FrameVirtuose.asp?IDLoja=3&I=6820
55	http://www.epocacosmeticos.com.br/	112	http://www.virtuose.com.br/Lojas/00000003/asp/FrameVirtuose.asp?IDLoja=3&I=8424
56	http://laffayette.com.br/index.php	113	
57	http://www.comucopia.com.br/	114	http://www.acrisoft.com.br/home

Endereço eletrônico dos Sites Avaliados		Endereço eletrônico dos Sites Avaliados	
115	http://www.virtuose.com.br/Lojas/00000003/asp/FrameVirtuose.asp?IDLoja=3&l=775	178	http://www.foxvideo.com.br/
116	http://www.akitemditudo.com.br/	179	http://www.tectotal.com.br/
117	http://www.virtuose.com.br/Lojas/00000003/asp/FrameVirtuose.asp?IDLoja=3&l=286&p=16604	180	http://www.starcineshop.com.br/
118	http://www.shopotica.com.br/	181	http://todaoferta.uol.com.br/
119	http://www.lojadabeleza.com.br/	182	http://www.sebodoratino.com.br/
120	http://www.navgps.com.br/	183	http://www.ciavirtualmix.com.br/loja/Default.asp
121	http://www.baseinfo.com.br/	184	http://www.xtremecenter.com.br/
122	http://www.superutilidades.com.br/	185	http://www.pipasan.com.br/
123	http://www.emporiomarchador.com.br/	186	http://loja.tray.com.br/loja/loja.php?loja=97088
124	http://www.abajouressantilarosa.com.br/	187	http://www.bagaqqio.com.br/
125	http://www.cinderelamoveis.com.br/	188	http://www.camerum.com.br/
126	http://www.mxmsex.com.br/	189	http://www.arcoeflecha.com.br/default.asp?ieparca=
127	http://www.compbras.com.br/	190	http://monracing.ckless.com/
128	http://www.fastfotos.com.br/	191	http://www.kabah.com.br/
129	http://www.astrus.com.br/	192	http://www.ecotrek.com.br/index.asp
130	http://www.parafinas.com.br/loja/index.asp	193	http://www.lojaskd.com.br/index.php
131	http://www.casadopabx.com.br/	194	http://meumoveldemadeira.com.br/index.html
132	http://www.shopesportes.com.br/	195	http://www.iacaremoveis.com.br/
133	http://www.bestgospel.com.br/	196	http://www.oregonscientific.com.br/
134	http://www.misteriosantigos.net/	197	http://www.oquedardepresente.com/
135	http://www.planetamegamix.com/	198	http://www.floresnaweb.com/
136	http://www.mercadoporto.com/	199	http://www.rumo.com.br/sistema/Home.asp?IDLoja=8104&Y=9395621327377&cch=
137	http://www.usbcompany.com.br/	200	http://www.floresdora.com.br/shop/index.asp?origem=Site
138	http://www.erotic toys.com.br/	201	http://www.grosetha.art.br/
139	http://www.meupabx.com.br/	202	http://www.amolare.com.br/
140	http://loja.universoeditorial.com.br/index.asp	203	http://www.pepper.com.br/novo/default.asp
141	http://www.15denovembro.com.br/	204	http://www.maserati.com/
142	http://www.lojaabril.com.br/	205	http://www.uniflores.com.br/
143	http://www.asiashop.com.br/sistema/home.asp?IDLoja=7773	206	http://www.maisquepresente.com.br/index.aspx
144	http://www.socravatas.com.br/	207	http://www.cestasmika.com/loja/index.php
145	http://www.vinhosdoc.com.br/	208	http://www.monarch.com.br/default.asp
146	http://www.rumo.com.br/sistema/home.asp?IDLoja=2982&Y=4210084096136	209	http://www.viainox.com/index.php?PID=12
147	http://www.tmart.com.br/sistema/home.asp?IDLoja=6713&1ST=1&Y=4210089186306	210	http://www.linuxmall.com.br/?category=camisetas
148	https://www.rumo.com.br/sistema/home.asp?IDLoja=7465&Y=4210090756358&cch=	211	http://www.marisa.com.br/scripts/capa.asp
149	http://www.crismel.com.br/	212	http://www.bandeira1.com.br/
150	http://www.armarinhos.com/	213	http://www.lgmimport.com/
151	http://www.mundodofutebol.com.br/	214	http://www.fastshox.com/sistema/home.asp?IDLoja=7334&1ST=1&Y=9396549308310
152	http://www.hopeshopping.com.br/sistema/home.asp?IDLoja=4477&Y=4210102396746	215	http://www.conexaoimport.com/
153	http://www.projtotamar.org.br/lojavirtual/	216	http://www4.ciashop.com.br/sportsoeste/
154	http://www.netbolsas.com.br/	217	http://www.calcadoonline.com.br/
155	http://www.macauly.com.br/index.asp	218	http://www.tennispoint.com.br/v2/?lang=pt_BR
156	http://www.clubedaostra.com.br/	219	http://www.zura.com.br/
157	http://www.dickjardim.com.br/	220	http://www.sportkids.com.br/?gclid=CNS26pinnZYCFR0RnQo dMiuA7Q
158	http://www.packandtrack.com.br/	221	http://www.prospin.com.br/default.aspx
159	http://www.amomeucarro.com.br/	222	http://www.moveyourself.com.br/
160	http://www.lingerie.com.br/	223	http://loja.cpapson.com.br/default.asp
161	http://www.cuecasonline.com.br/	224	http://www.commcen.com.br/index.php
162	http://www.livrariatapioca.net/	225	http://www.portalgenial.com/loja/
163	http://www.shopsmile.com.br/	226	http://www.lebes.com.br/
164	http://www.comercialmoema.com.br/	227	http://www.piauiimport.com/index.php
165	http://www.3dvirtua.com/	228	http://www.cdmedia.com/default.asp
166	http://www.553exact.com.br/	229	https://www.playplaybrinquedos.com/default.asp
167	http://www.abcloja.com.br/	230	http://saciperere.com.br/
168	http://videolar.com/	231	http://www.balloonbrinquedos.com.br/loja/
169	http://www.dvdworld.com.br/dvdworld.htm	232	http://www.supercanguru.com.br/
170	http://www.arenadvd.com.br/	233	http://www.oprojetista.com.br/
171	http://www.allcenter.com.br/	234	http://www.casacruz.com.br/pg_inicial.cfm
172	http://www.unicabr.com.br/index.html	235	http://www.ecanetas.com.br/default.asp
173	http://www.livrariasaraiva.com.br/	236	http://www.seny.com.br/
174	http://www.colombo.com.br/	237	http://www.vetorialinfoja.com/default.php
175	http://www.fastshop.com.br/default.aspx	238	http://www.photoink.com.br/default.asp
176	http://www.dvdintemacional.com.br/	239	http://www.lojaprolar.com.br/index.php
177	http://www.apoderosa.com.br/	240	http://www.imashop.com.br/sistema/home.asp?IDLoja=6599

Endereço eletrônico dos Sites Avaliados		Endereço eletrônico dos Sites Avaliados	
241	http://www.inforoffice.com.br/	302	http://www.rhinoceronte.com.br/
242	http://www.amorcomprazer.com.br/loja/default.asp	303	http://www.cremonesi.com.br/
243	http://www.posthaus.com.br/	304	http://www.brasutil.com/?lang=pt_BR
244	http://www.rockstore.com.br/store/comersus_index.asp	305	http://www.amoramiura.com/
245	http://www.sennastore.com.br/	306	https://www.baladainformatica.com.br/
246	http://www.disantinni.com.br/index.aspx	307	http://www.babelshop.com.br
247	https://lojaqx1.websiteseguro.com/index.php	308	http://www.allbags.com.br/Home4.aspx
248	http://obashop.com/v4/default.asp	309	http://www.lojasscintos.com.br/
249	http://www.gospelgoods.com.br/	310	http://www.produtosvirtuais.com/
250	http://obashop.com/v4/default.asp	311	http://www.fisiostore.com.br/
251	http://www.amafali.com.br/vitrine.php	312	http://www.oticaov.com.br/
252	http://www.sefer.com.br/	313	http://www.hendrixmusic.com.br/
253	http://www.lojadocepaixao.com/Default.asp	314	https://www.kabum.com.br/
254	http://www.intemetible.com.br/lojavirtual/	315	http://www.idepot.com.br/canais/hp/index.asp
255	http://www.mercadooriental.com.br/	316	http://www.palmetal.com.br
256	http://www.cpb.com.br/	317	http://www.strawbervynet.com/
257	http://www.bitcao.com.br/	318	http://www.asvip.com/
258	http://www.induspar.com/	319	http://lx1.letti.com.br/b2c/frutodearte/
259	http://www.accesinternacional.com.br/?lang=pt_BR	320	http://www.casadaarte.com.br/lista_grupo.php?grupo=306
260	http://www.polishop.com.br/	321	http://www.pinceldarte.com/
261	http://www.rosacalida.com.br/index.php?cPath=17	322	http://www.nicearts.com.br/
262	http://www.erosmania.com.br/	323	http://www.brinquedoslaura.com.br/produtos.aspx?d=peleucias&id=23
263	http://www.megaerotica.com.br/	324	http://www.lojaunicef.org.br/?gclid=CJOUzouyZcCFQazsgodRzC5-A
264	http://www.sexxyshop.com.br/welcome.php?pid=120543	325	http://www.floresechocolates.com.br/
265	https://www.sampaerotica.com.br/loja/index.php	326	http://www.toymagazine.com.br/index.php
266	http://www.amorcomprazer.com.br/loja/default.asp	327	http://www.edfilmadoras.com.br/index.asp
267	http://www.rhadija.com/loja/default.asp	328	http://www.ultramarket.com.br/
268	http://www.csul.com.br/default.asp	329	http://www.atelnetronicos.com.br/loja/
269	http://www.dogcharme.com.br/?lang=pt_BR	330	http://www.ciashop.com.br/gameone/
270	https://www.digitalup-lojavirtual.com.br/	331	http://www.lojasdavi.com.br/loja/
271	http://www.manole.com.br/livros.php?id=1743	332	http://www.cdmedia.com/default.asp
272	http://www.megaliros.com.br/home	333	http://www.pcfioripa.com.br/loja/index.php
273	http://www.plugme.com.br/	334	http://www.shazamshop.com.br/default.asp
274	http://www.traca.com.br/	335	http://www.realishop.com.br/website/index.php
275	http://br.gojaba.com/	336	http://www.flybox.com.br/
276	https://www.omdl.com.br/?gclid=CITQzKaynZYCFRkQnQod7XFBW	337	http://www.etrionics.com.br/default.asp
277	http://www.livrosnoro.com/	338	http://www.ishop21.com.br/home.aspx
278	http://www.johnsomers.com.br/	339	http://www.netinformatica.org/default.asp
279	http://www.livroteca.com.br/index.php	340	http://www.gpi.com.br/default.aspx
280	http://www.chocolatecaseiro.com.br/loja/index.php	341	http://www.sonystyle.com.br/
281	http://www.mercurysshop.com.br/	342	http://www.ciashop.com.br/starhouse/default.asp?template_id=62&partner_id=8
282	http://www.brazilvshop.com.br/	343	http://www.beephoto.com.br/default.asp
283	http://www.dedobrinquedo.com.br/	344	http://www.senvy.com.br/
284	http://www.brasilmultiart.com.br/vitrine.aspx	345	http://www.mosquiteirobrasil.com.br/
285	http://www.ponta destock.com.br/	346	http://www.bebeweb.com.br/
286	http://www.arteponto1.com.br/?lang=pt_BR	347	http://www.angelinababy.com.br/
287	http://www.mixdasessencias.com.br/	348	http://www.abcmusical.com.br/
288	http://www.rumo.com.br/sistema/home.asp?IDLoja=1661	349	http://www.calcadoonline.com.br/
289	http://www.ecomarine.com.br/	350	http://www.posthaus.com.br/lojas/posthaus?acao=home
290	http://expand.americanas.com.br/controller.asp?origem=expand	351	http://www.disantinni.com.br/index.aspx
291	http://www.wineexpress.com.br/	352	http://www.limaparafusos.com.br/loja/index2.php
292	http://www.winestore.com.br/principal/	353	http://www.varejaonet.com.br/
293	http://www.emporiofrances.com/index2.php	354	http://www.planetadeagostini.com.br
294	http://www.agataweb.com.br/	355	http://www.petfast.com.br/sistema/home.asp?IDLoja=4783&Y=2929148138271
295	http://www.invinovita.com.br/?gclid=CJl-kuXL25YCFQETGgod4jhT2q	356	http://www.petcentemarginal.com.br/
296	http://plixx.com.br/loja/	357	http://www.virtualpet.com.br/loja/default.asp
297	http://www3.ciashop.com.br/avicente/default.asp?template_id=60&partner_id=8	358	http://www.onofre.com.br/onofre/home.aspx
298	http://www.portcasa.com.br/	359	http://www.drogaraia.com.br/RaiaEcommWeb/
299	http://www.ciadsoftware.com.br/	360	http://www.derma doctor.com.br/
300	http://www.compredachina.com/	361	http://www.ultrafarma.com.br/local.aspx
301	http://www.zenittecnologia.com.br/		

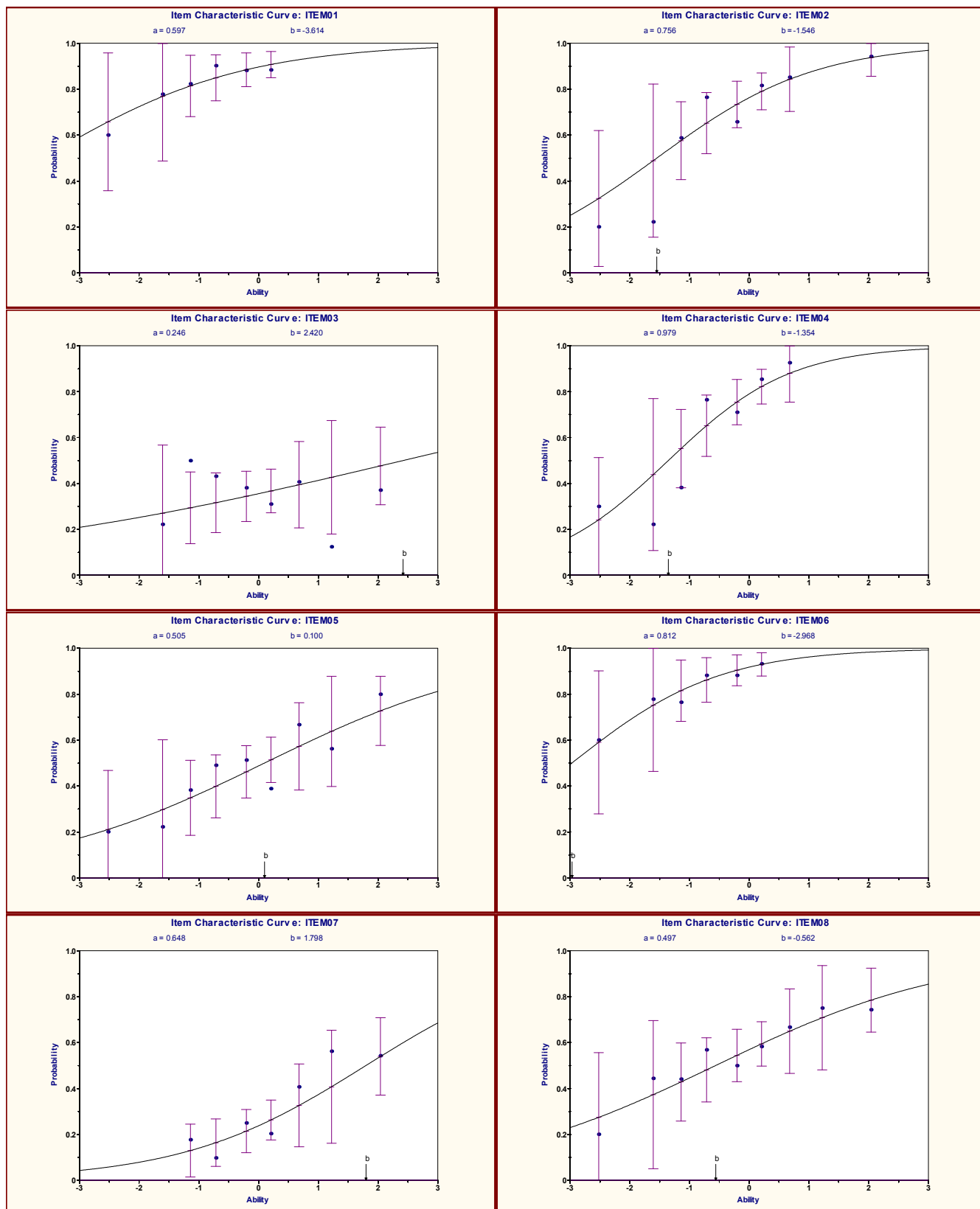
APÊNDICE B - Conjunto de itens iniciais

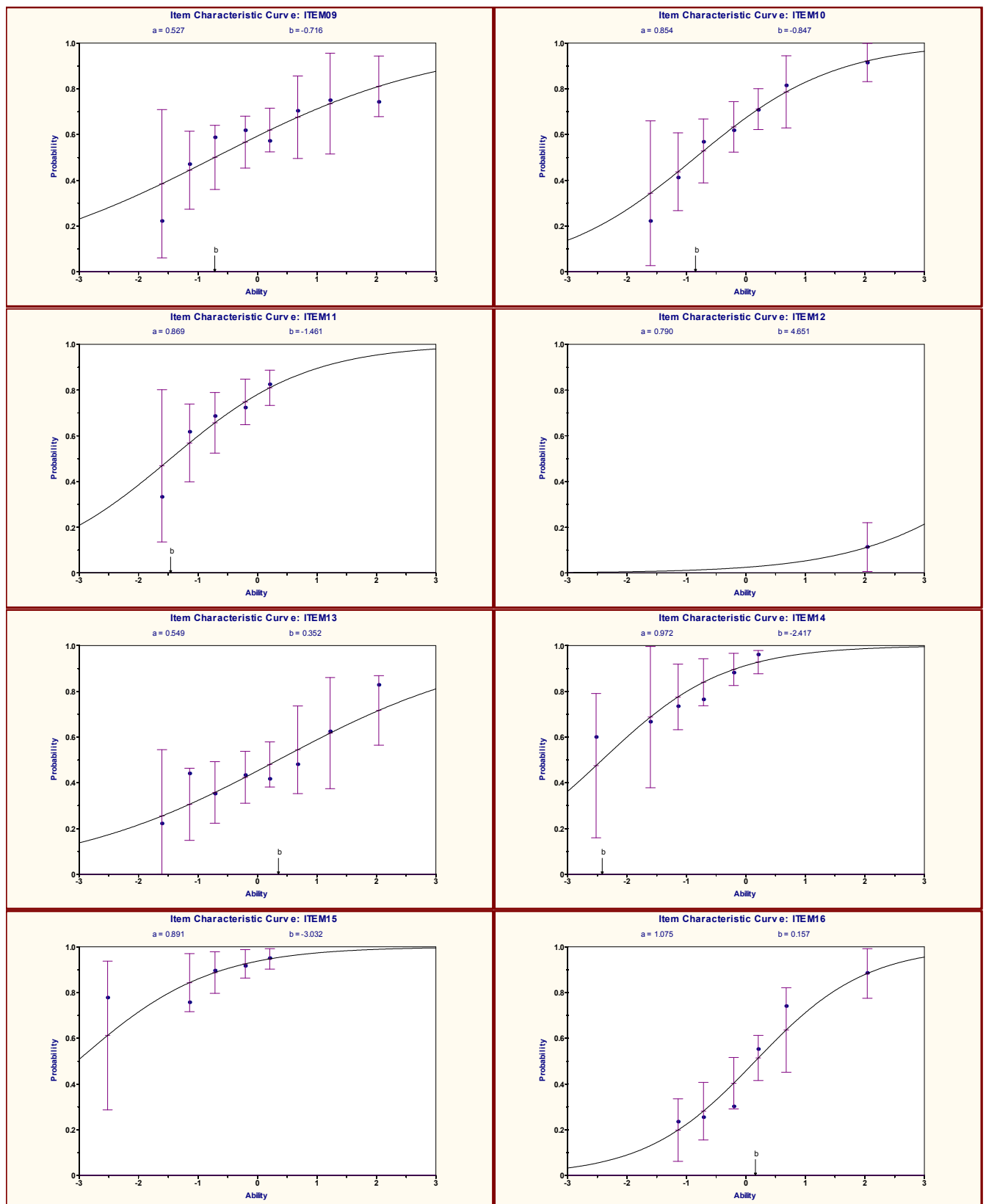
Item	Descrição
1	Todas as páginas possuem os mesmos <i>layouts</i> e exibem ao usuário as mesmas características?
2	O logotipo da empresa está no canto superior esquerdo em todas as páginas do site?
3	Existe um link de um único clique que conduz à página inicial?
4	O botão “voltar” funciona em todas as janelas?
5	O site utiliza um único ícone para uma determinada função?
6	Existe uma orientação ao usuário quanto ao restante do site?
7	Os textos possuem um bom contraste?
8	A página inicial deixa claro o que o site faz?
9	A página inicial possui menos de 2,3 páginas de rolagem?
10	Os links já visitados mudam de cor?
11	Ao clicar em um link ou em um botão não é aberta uma outra janela?
12	Quando há necessidade de abrir outra janela está não aparece maximizada?
13	Ao abrir a página inicial, não há abertura de janelas pop-up?
14	As palavras aparentemente clicáveis são de fato clicáveis ?
15	As palavras clicáveis apresentam uma forma diferenciada quando são selecionadas?
16	Se houver vídeo clipe, este tem duração de menos de um minuto?
17	O layout da página é adaptável ao tamanho da janela, pelo menos horizontalmente?
18	Todas as palavras em cores diferenciadas e/ou sublinhada são clicáveis?
19	Quando há rolagem, não existe elementos de design (na tela inicial) que pareçam com marcadores de final de página?
20	As opções principais do site estão visíveis? Ou seja, não há necessidade de passar o mouse sobre uma área gráfica para ver opções ocultas?
21	Os menus em cascata possuem no máximo dois níveis?
22	Todos os campos e mostradores de dados possuem rótulos identificativos?
23	O site possui a opção “ajuda” ?
24	Letras em maiúsculo se restringem a títulos ou pequenas frases clicáveis, que não ultrapassam uma linha?
25	O tamanho da fonte dos textos é de dez pontos ou mais?
26	Na área principal do site, existem textos com no máximo quatro cores diferentes?
27	Na área principal do site existem textos com no máximo três fontes diferentes?
28	Os títulos principais têm um tamanho maior e mais destacado que o corpo do texto?
29	Os títulos estão alinhados à esquerda?
30	As características dos produtos, quando maiores que três, são listadas verticalmente ao invés de horizontais e contínuas?
31	Os itens das listas possuem um paralelismo sintático?
32	O preço de um produto consta ao lado da imagem ou do link do produto (ou seja, não há necessidade de clicar no produto para obter informações de preço. Mesmo que haja variação de lojas, é bom mostrar uma estimativa)?
33	É possível saber os custos totais antes de fazer cadastro?
34	A maioria dos produtos possui informações sobre eles?

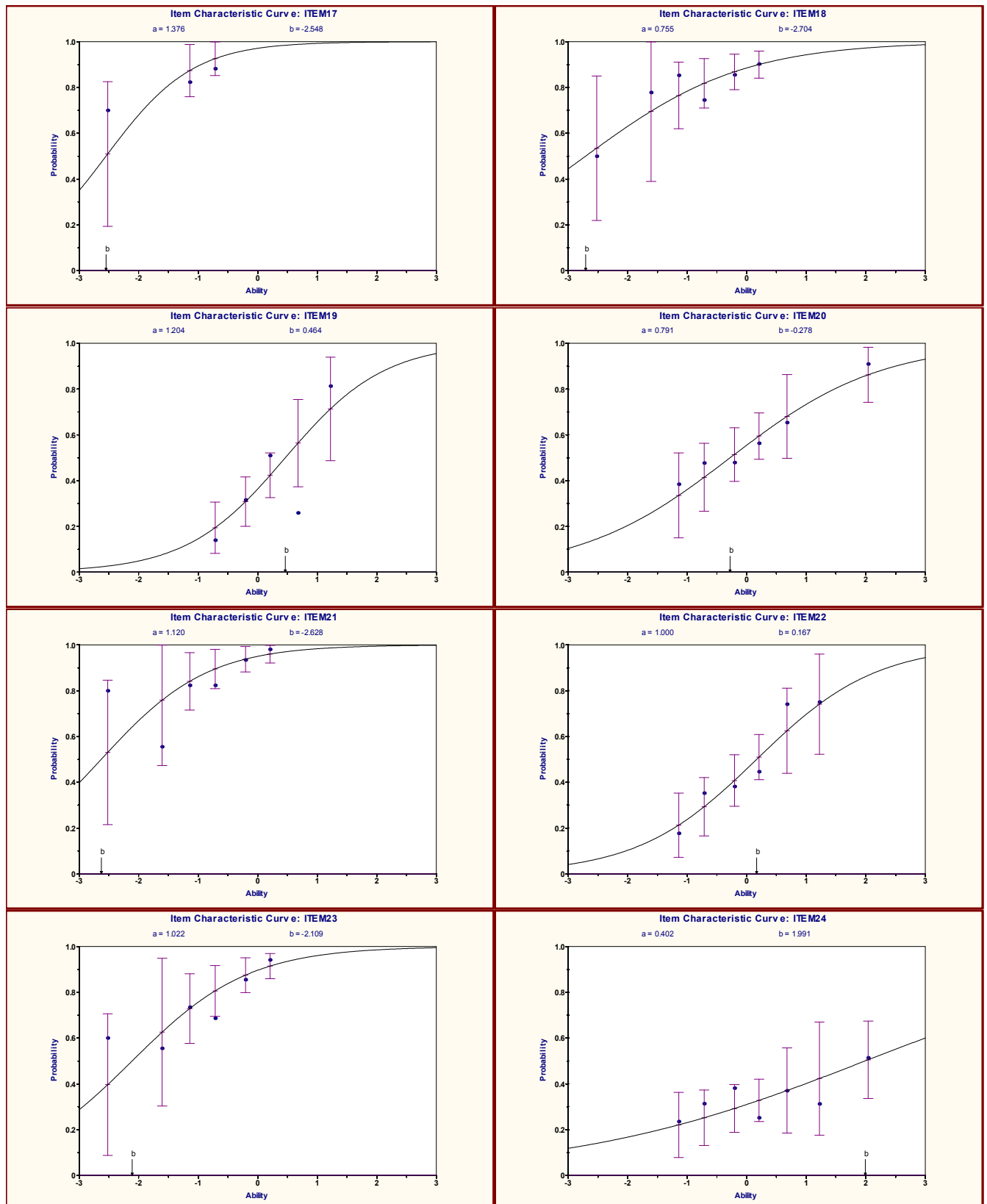
Item	Descrição
35	É possível ampliar as fotos dos produtos para visualizar detalhes?
36	Em produtos em que existam mais de uma perspectiva, é possível visualizar todas as perspectivas?
37	Os campos destinados a formulários de preenchimento pelo usuário estão agrupados linearmente, evitando espaços desnecessários?
38	O botão de finalização de compra está no final da lista?
39	Existem espaços em branco?
40	É possível encontrar informações sobre a empresa (endereço físico, executivos etc)?
41	A barra de rolagem está a direita possui um formato padrão?
42	Todas as páginas possuem um campo de busca padronizado?
43	Todos os produtos que são mostrados no site podem ser comprados?
44	É possível visualizar o custo total do produto (custo do produto mais impostos mais transporte) antes de fazer qualquer cadastro?
45	Os resultados de busca permitem classificação por outros critérios, além de custo?
46	As áreas ativas de um link estão em área visível próxima?
47	É possível efetivar uma compra sem realizar um cadastro (que inclua login e senha)?
48	O site possui opção de acesso em outras línguas?
49	Quando existir rolagem vertical, isto é perceptível na primeira página ?
50	Todos os textos são perfeitamente legíveis?
51	A aparência dos links acessados é modificada? O sistema fornece um histórico dos comandos entrados pelo usuário durante uma sessão de trabalho?
52	Ao ser passado o mouse sobre algum objeto este é realçado?
53	Toda ação de entrada de dados fornece informações de ajuda?
54	No preenchimento de um formulário, as opções que não são válidas ou não estão disponíveis estão visivelmente desativadas, prevenindo erros?
55	No preenchimento de um formulário, é informada a forma de preenchimento?
56	Os títulos de telas, janelas e caixas de diálogo estão no alto, centrados ou justificados à esquerda?
57	As unidades para a entrada ou apresentação de dados métricos ou financeiros encontram-se descritas na tela?
58	Listas longas apresentam indicadores de continuação, de quantidade de itens e de páginas?
59	Os botões que comandam a apresentação de caixas de diálogo apresentam em seus rótulos o sinal "... " como indicador da continuidade do diálogo?
60	Nas caixas de mensagens de erro, o botão de comando "AJUDA" está sempre presente?
61	O espaço de apresentação está diagramado em pequenas zonas funcionais?
62	A disposição dos objetos de interação de uma caixa de diálogo segue uma ordem lógica?
63	Os grupos de botões de comando estão dispostos em coluna e à direita, ou em linha e abaixo dos objetos aos quais estão associados?
64	Os dados obrigatórios são diferenciados dos dados opcionais de forma visualmente clara?
65	O sistema fornece "feedback" para todas as ações do usuário?
66	Os parágrafos de texto são separados?
67	A altura mínima dos mostradores de texto é de 4 linhas?
68	Os rótulos de campos começam com uma letra maiúscula e as letras restantes são minúsculas?

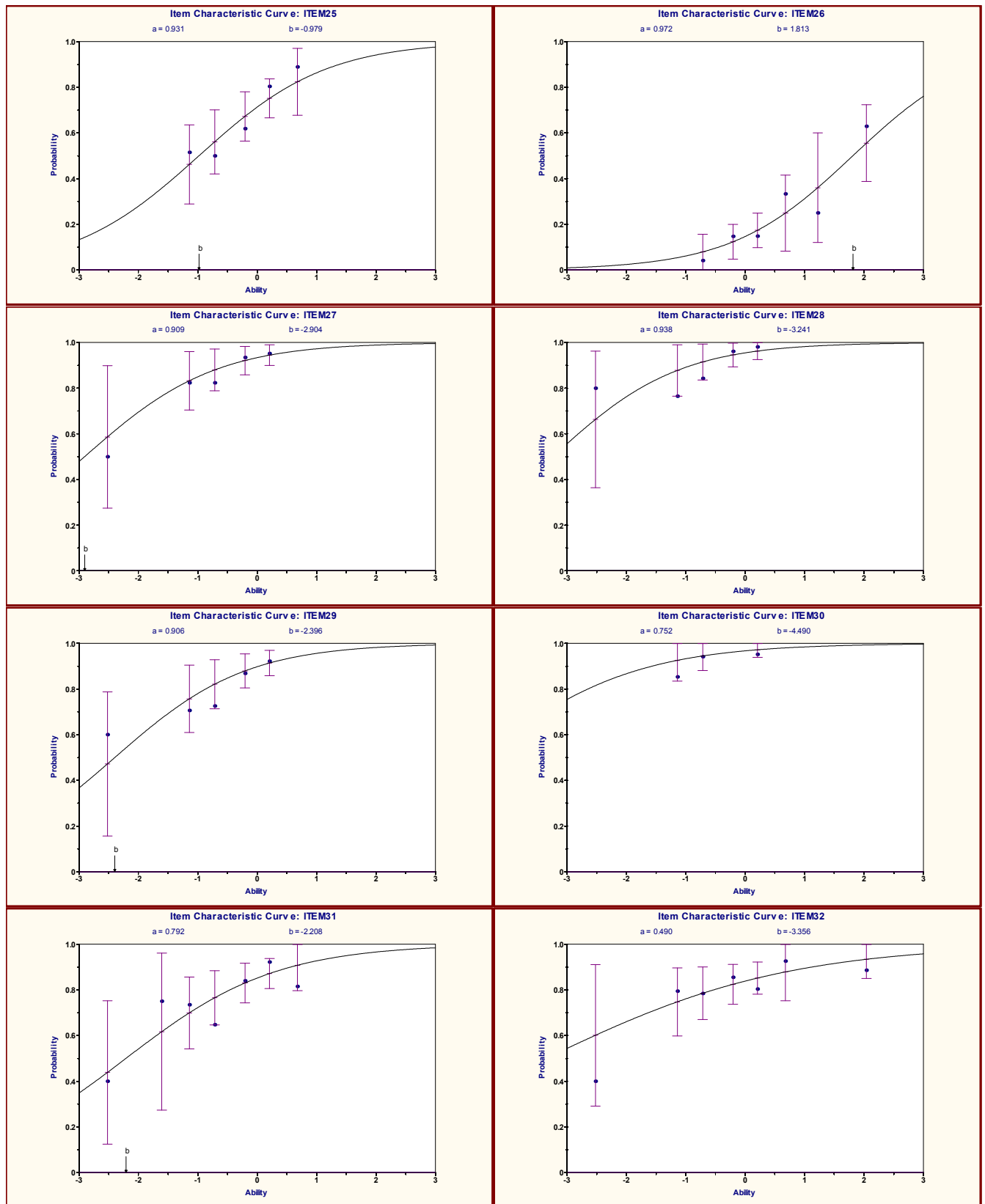
Item	Descrição
69	O sistema emite sinais sonoros quando ocorrem problemas na entrada de dados?
70	O sistema solicita confirmação dupla de ações que podem gerar perdas de dados e/ ou resultados catastróficos?
71	As mensagens de erro estão isentas de abreviaturas e/ ou códigos gerados pelo sistema operacional?
72	Qualquer ação do usuário pode ser revertida através da opção DESFAZER?
73	Os menus em cascata possuem uma área de arraste com espessura > X cm ?

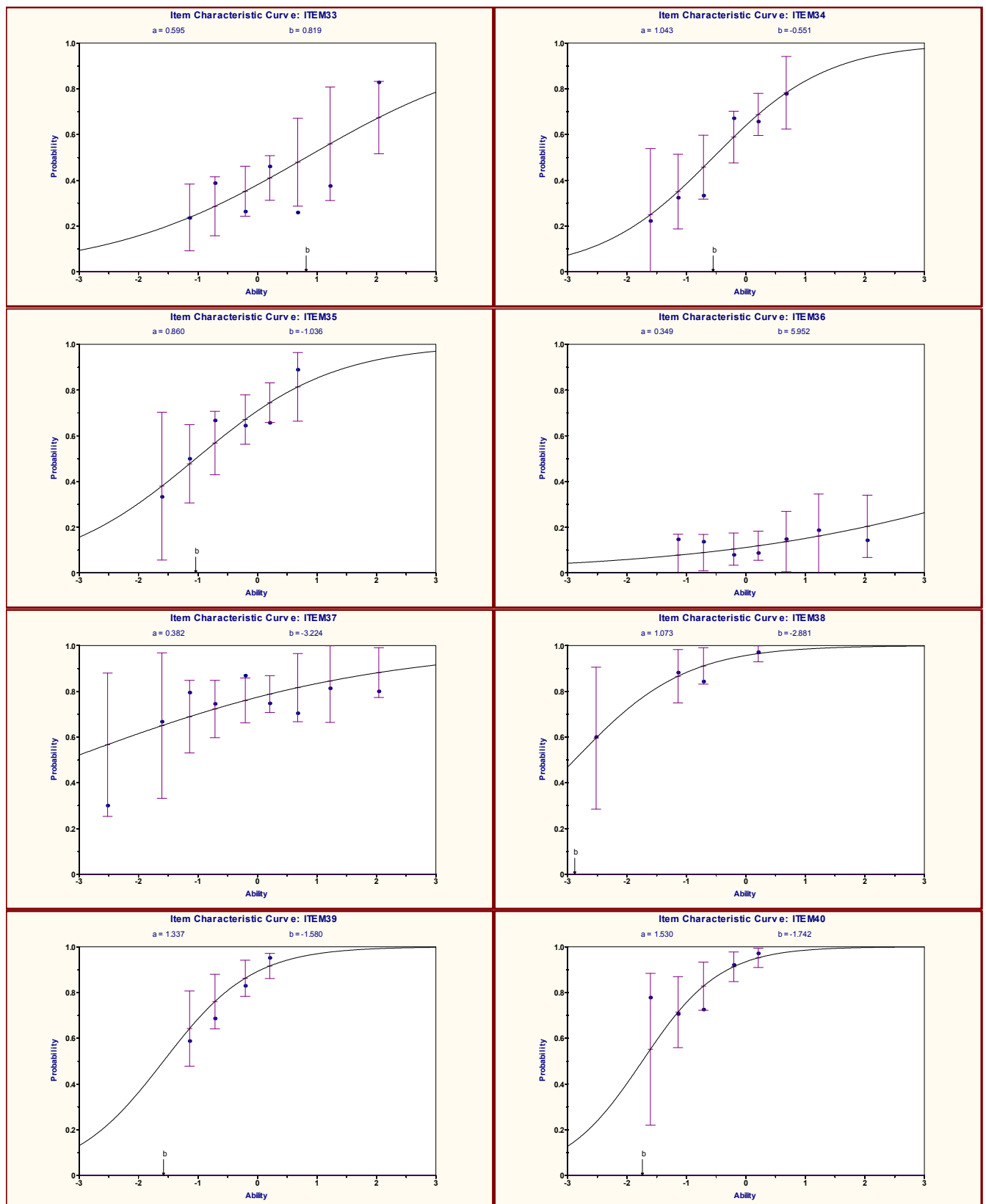
APÊNDICE C – Curvas Características dos Itens

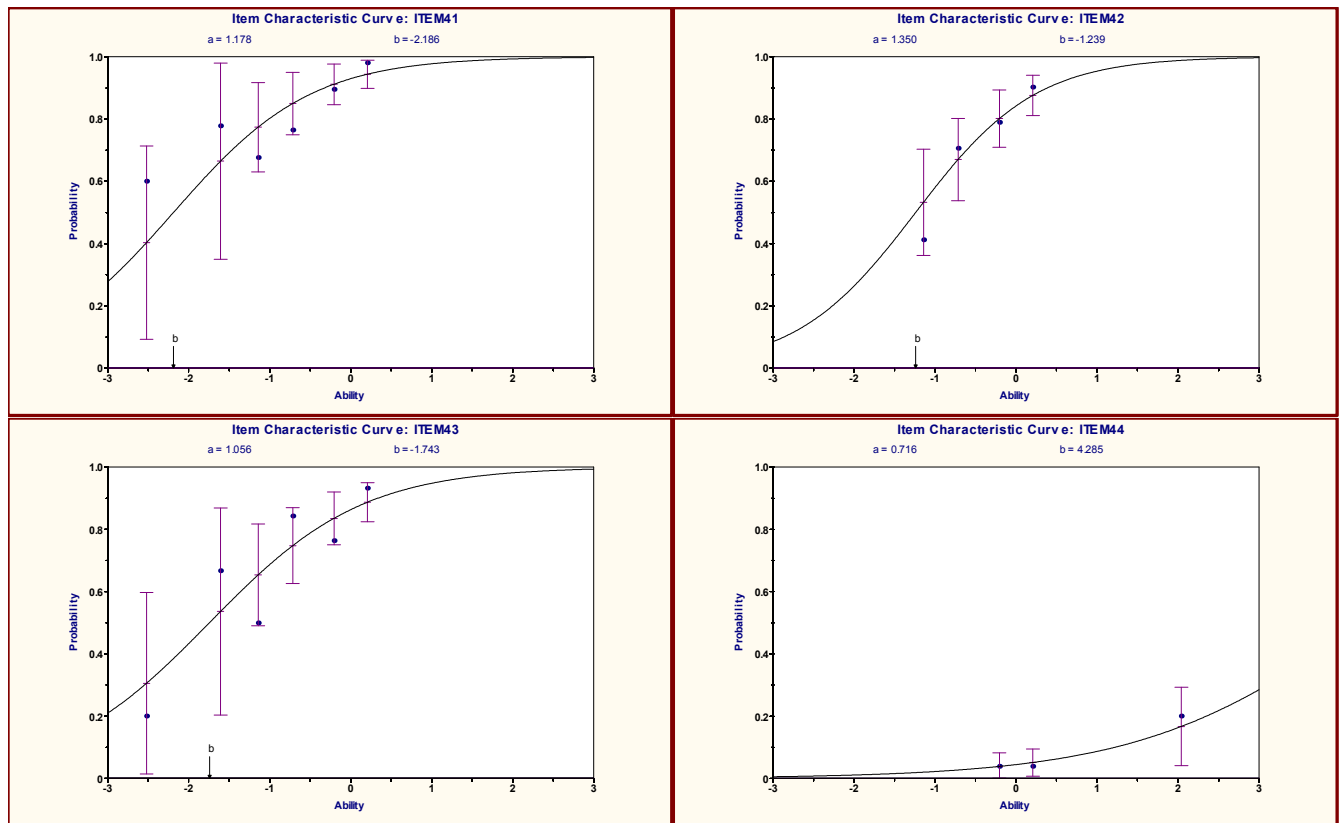




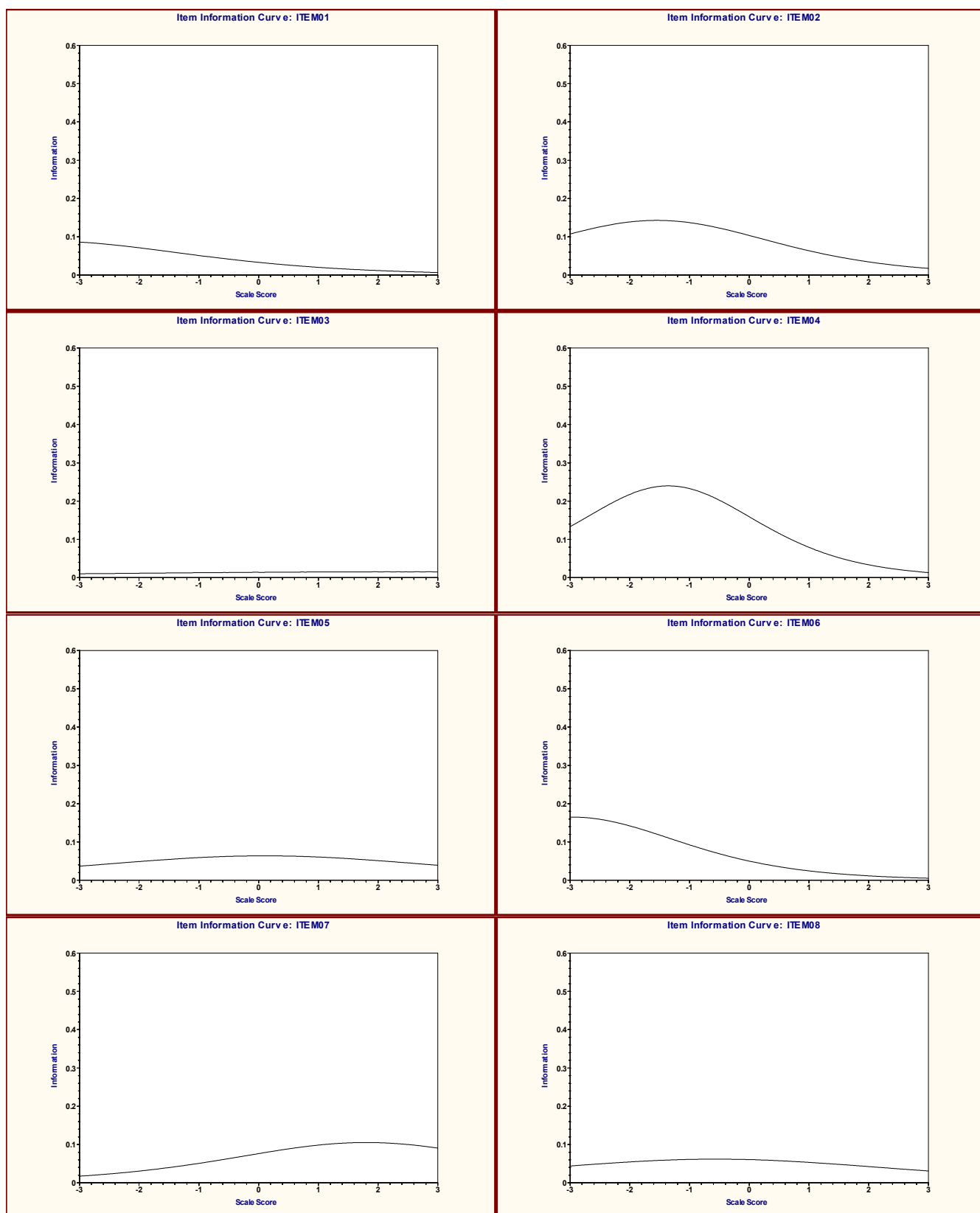


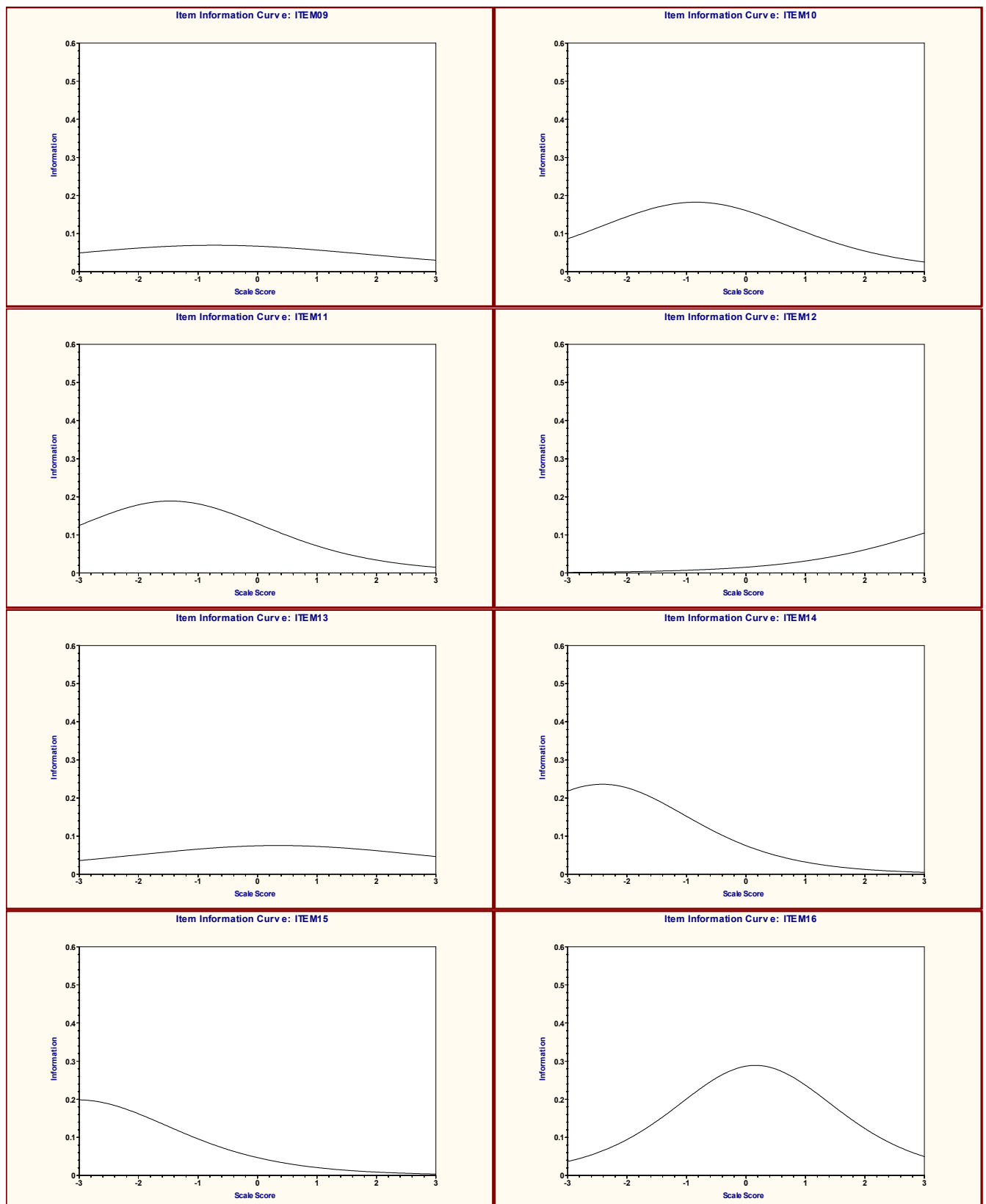


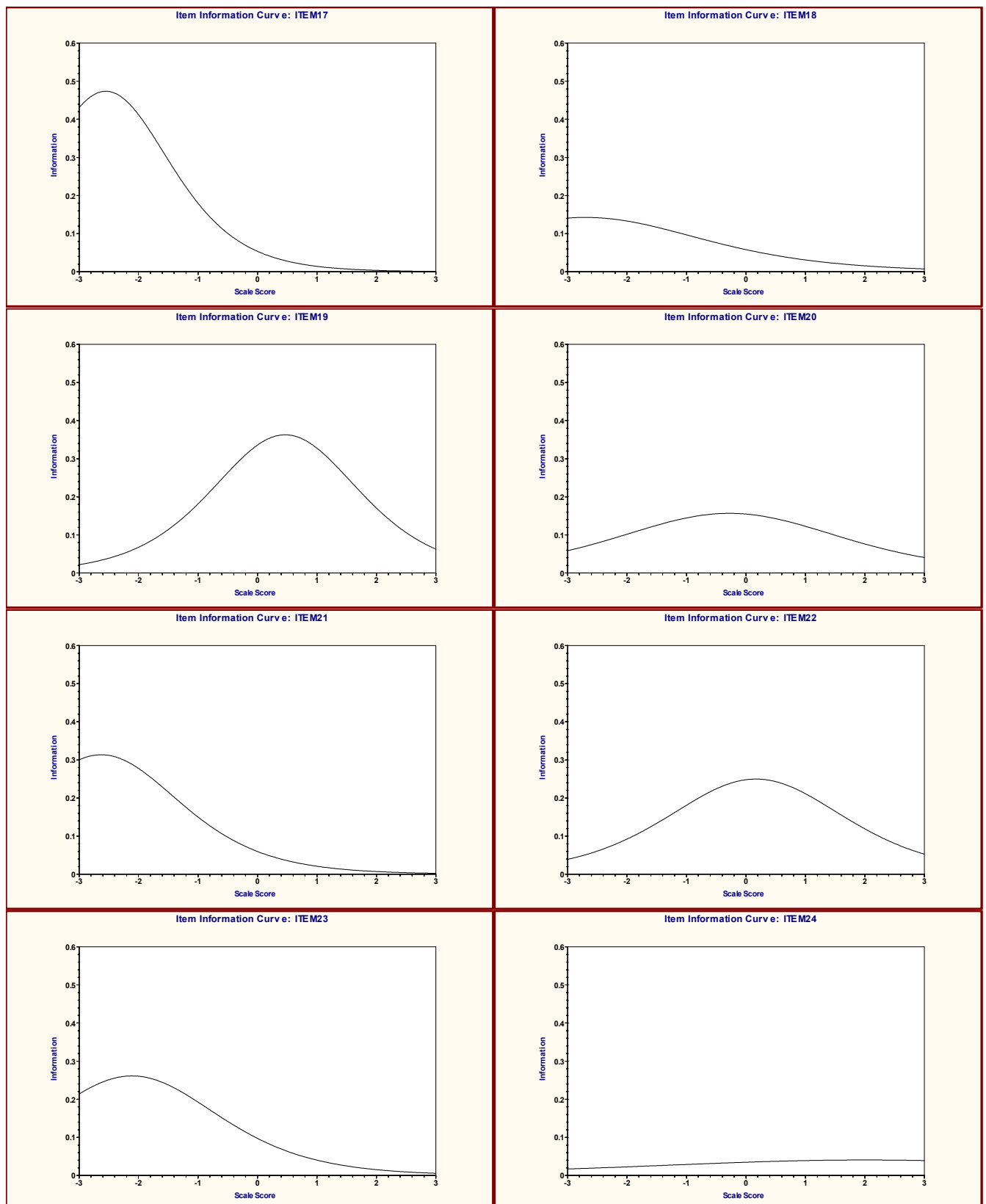


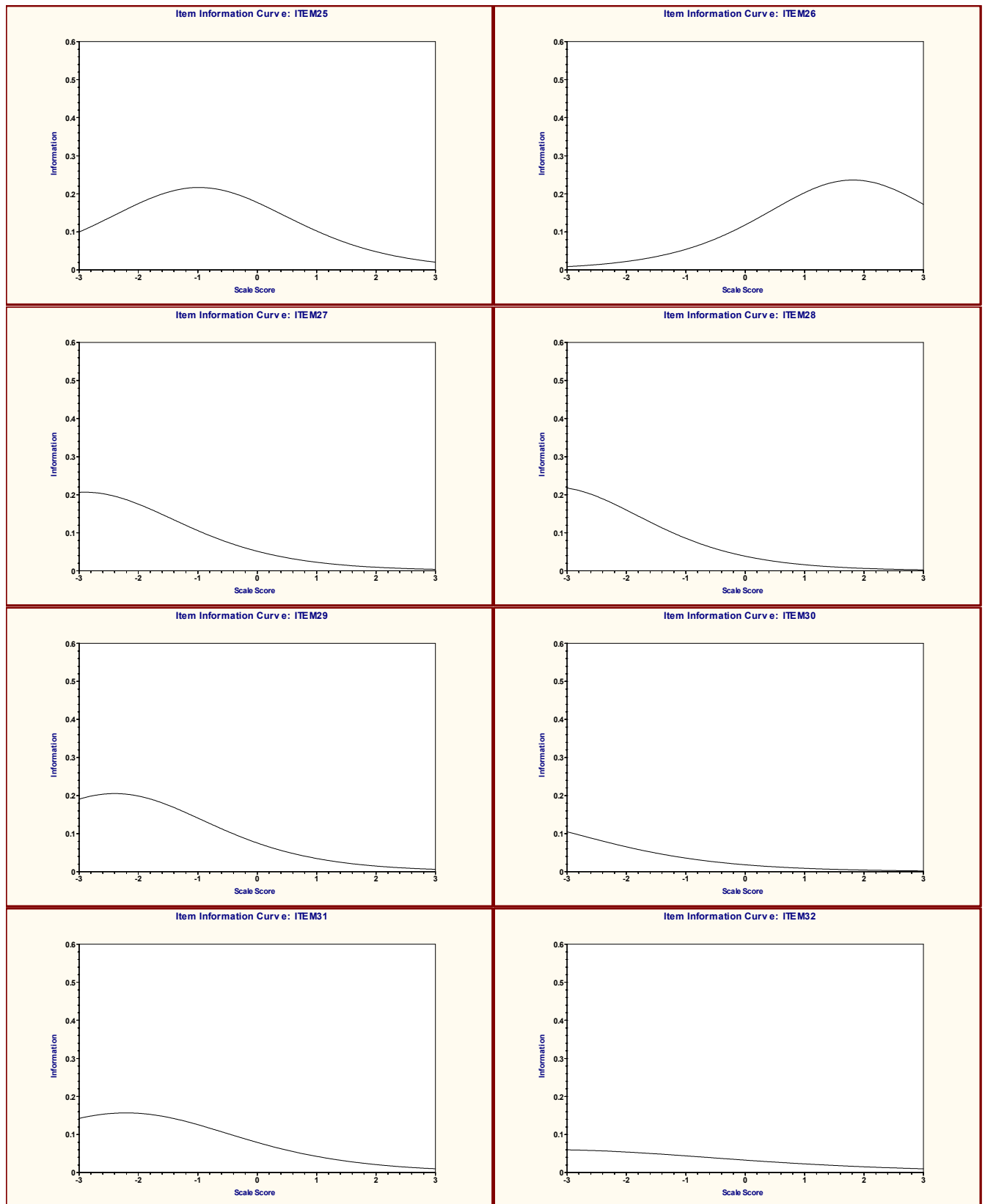


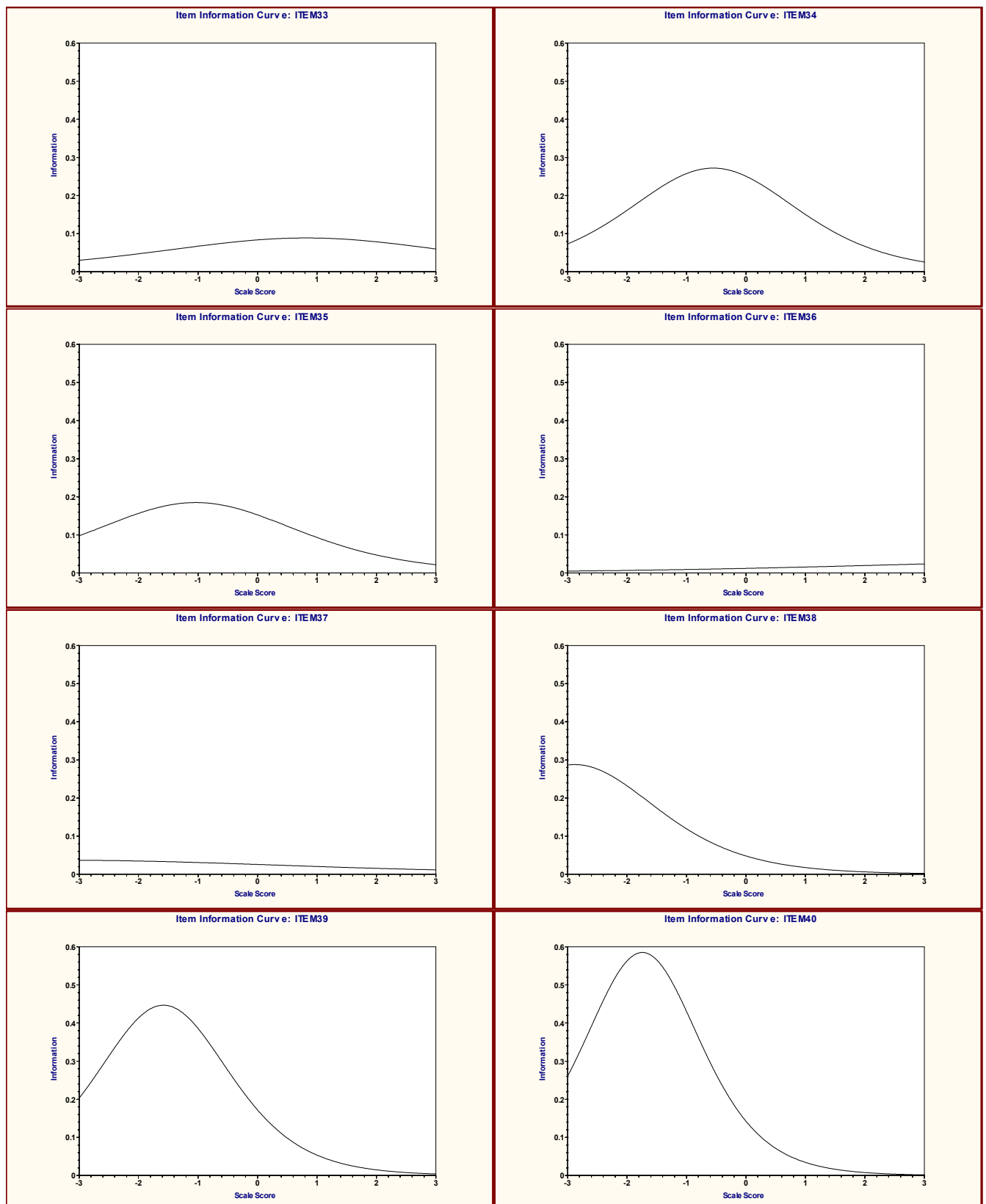
APÊNDICE D – Curva de Informação dos Itens

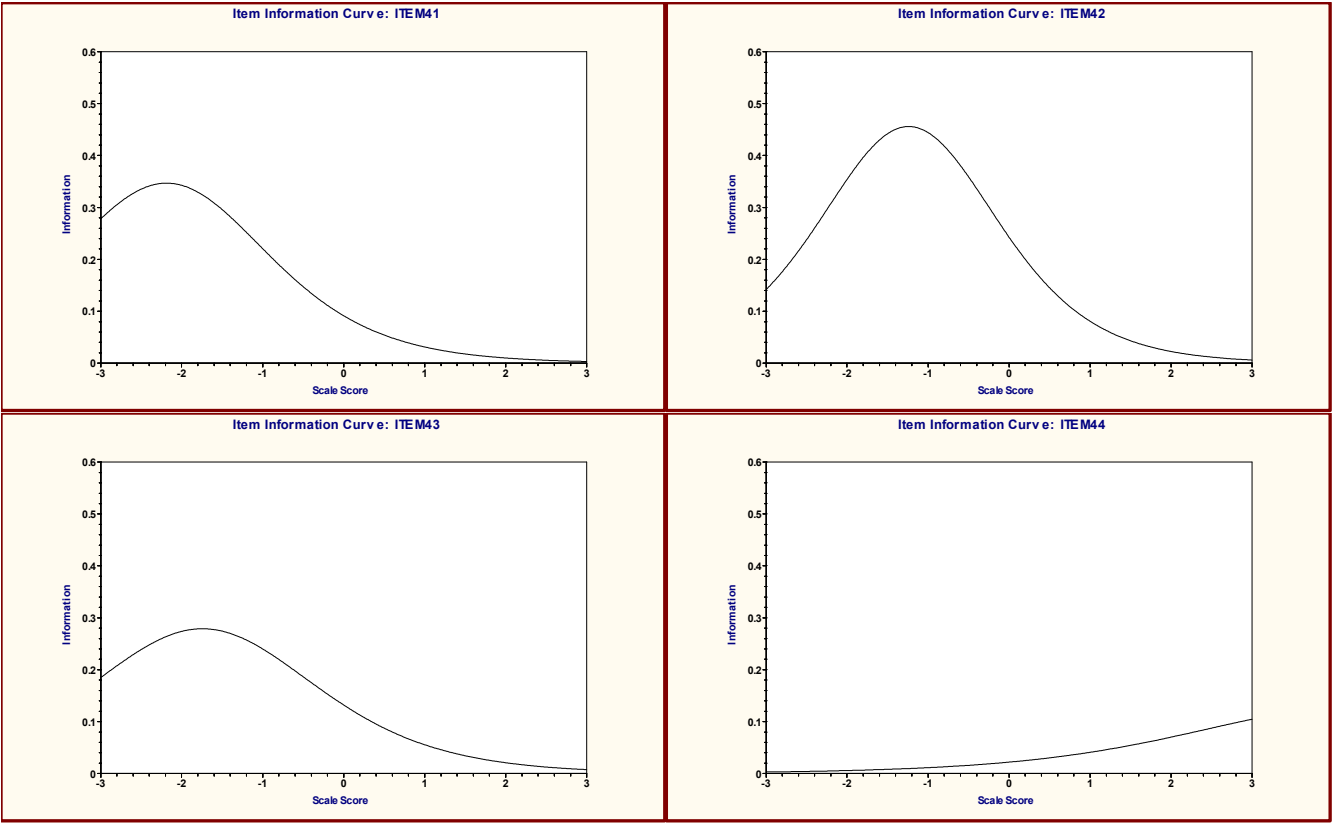












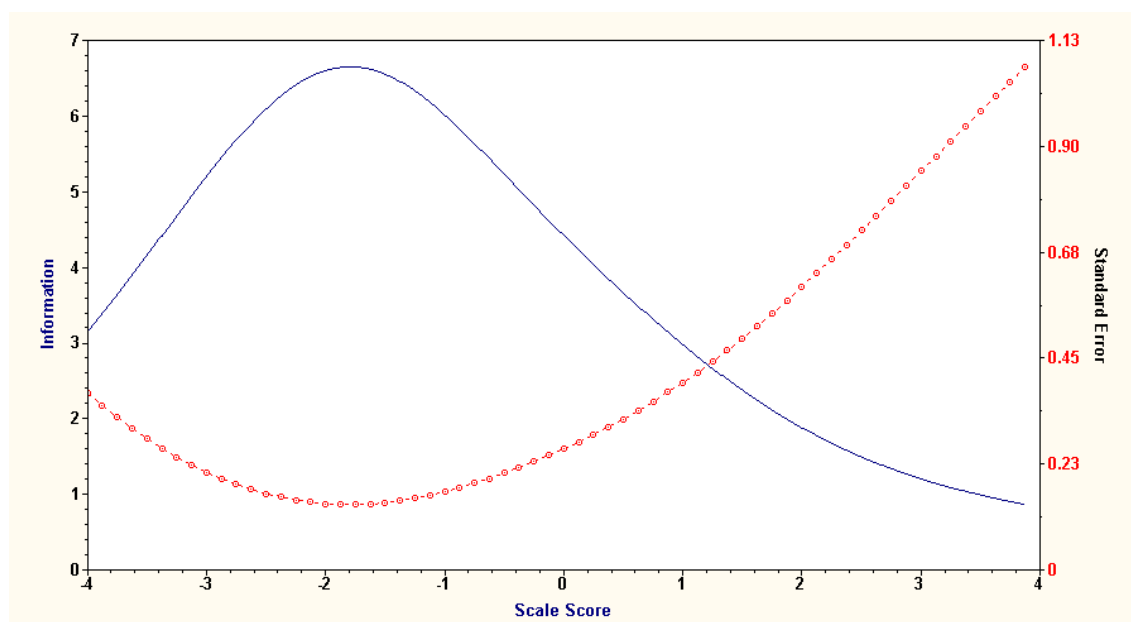
APÊNDICE E – Estimação dos parâmetros após a retirada de alguns itens.

Item	44 itens				43 itens				42 itens				41 itens				40 itens			
	Par. a	erro	Par. b	erro	Par. a	erro	Par. b	erro	Par. a	erro	Par. b	erro	Par. a	erro	Par. b	erro	Par. a	erro	Par. b	erro
01	0,597	0,155	-3,614	0,897	0,584	0,152	-3,685	0,925	0,583	0,152	-3,691	0,926	0,578	0,151	-3,721	0,937	0,583	0,153	-3,688	0,927
02	0,756	0,153	-1,546	0,311	0,762	0,153	-1,536	0,308	0,757	0,152	-1,544	0,310	0,765	0,152	-1,532	0,304	0,773	0,154	-1,519	0,300
03	0,246	0,067	2,420	0,798	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
04	0,979	0,175	-1,354	0,229	0,966	0,174	-1,369	0,233	0,967	0,173	-1,368	0,232	0,955	0,170	-1,381	0,234	0,963	0,171	-1,373	0,231
05	0,505	0,115	0,100	0,216	0,490	0,112	0,103	0,222	0,486	0,111	0,104	0,224	0,474	0,109	0,106	0,229	0,474	0,109	0,106	0,229
06	0,812	0,193	-2,968	0,613	0,825	0,195	-2,931	0,599	0,822	0,194	-2,941	0,601	0,822	0,194	-2,940	0,603	0,821	0,193	-2,942	0,602
07	0,648	0,127	1,798	0,369	0,646	0,126	1,804	0,371	0,639	0,125	1,821	0,376	0,640	0,125	1,818	0,375	0,637	0,125	1,825	0,378
08	0,497	0,110	-0,562	0,246	0,490	0,109	-0,569	0,249	0,491	0,109	-0,568	0,248	0,492	0,109	-0,567	0,248	0,492	0,110	-0,567	0,247
09	0,527	0,115	-0,716	0,255	0,513	0,113	-0,733	0,262	0,517	0,114	-0,729	0,260	0,503	0,112	-0,746	0,268	0,502	0,112	-0,748	0,268
10	0,854	0,162	-0,847	0,189	0,853	0,162	-0,848	0,189	0,858	0,162	-0,845	0,188	0,852	0,161	-0,850	0,189	0,845	0,160	-0,856	0,190
11	0,869	0,159	-1,461	0,268	0,860	0,157	-1,473	0,270	0,860	0,156	-1,473	0,269	0,847	0,152	-1,491	0,272	0,853	0,154	-1,484	0,270
12	0,790	0,212	4,651	1,134	0,785	0,211	4,678	1,145	0,773	0,208	4,736	1,168	0,781	0,208	4,697	1,142	0,775	0,210	4,725	1,171
13	0,549	0,119	0,352	0,216	0,544	0,118	0,355	0,217	0,545	0,118	0,354	0,217	0,553	0,119	0,349	0,214	0,557	0,119	0,347	0,213
14	0,972	0,225	-2,417	0,446	0,990	0,228	-2,384	0,434	0,988	0,228	-2,387	0,436	1,008	0,232	-2,352	0,424	1,022	0,235	-2,329	0,417
15	0,891	0,228	-3,032	0,641	0,901	0,229	-3,003	0,627	0,898	0,228	-3,013	0,630	0,906	0,229	-2,991	0,622	0,912	0,231	-2,973	0,618
16	1,075	0,178	0,157	0,114	1,066	0,178	0,158	0,115	1,063	0,178	0,159	0,116	1,074	0,179	0,157	0,115	1,066	0,177	0,158	0,115
17	1,376	0,346	-2,548	0,401	1,384	0,352	-2,538	0,403	1,384	0,354	-2,538	0,404	1,403	0,362	-2,515	0,402	1,437	0,371	-2,478	0,388
18	0,755	0,183	-2,704	0,592	0,773	0,186	-2,654	0,572	0,785	0,188	-2,619	0,559	0,785	0,188	-2,621	0,561	0,786	0,189	-2,616	0,560
19	1,204	0,228	0,464	0,123	1,227	0,231	0,459	0,121	1,238	0,232	0,456	0,120	1,268	0,238	0,448	0,117	1,271	0,237	0,448	0,117
20	0,791	0,159	-0,278	0,158	0,784	0,158	-0,280	0,160	0,773	0,156	-0,283	0,162	0,763	0,154	-0,287	0,164	0,773	0,156	-0,284	0,162
21	1,120	0,283	-2,628	0,480	1,136	0,287	-2,600	0,471	1,149	0,288	-2,580	0,462	1,170	0,294	-2,547	0,452	1,173	0,294	-2,543	0,451
22	1,000	0,179	0,167	0,123	0,980	0,176	0,170	0,125	0,983	0,176	0,170	0,125	0,990	0,177	0,168	0,124	1,004	0,177	0,166	0,123
23	1,022	0,226	-2,109	0,367	1,036	0,228	-2,088	0,360	1,027	0,226	-2,103	0,364	1,038	0,228	-2,086	0,359	1,036	0,227	-2,089	0,359
24	0,402	0,100	1,991	0,551	0,405	0,100	1,977	0,545	0,405	0,101	1,977	0,546	0,410	0,101	1,955	0,538	-	-	-	-
25	0,931	0,165	-0,979	0,186	0,957	0,168	-0,960	0,180	0,956	0,168	-0,960	0,180	0,971	0,169	-0,950	0,176	0,969	0,169	-0,952	0,176
26	0,972	0,179	1,813	0,294	0,982	0,179	1,800	0,290	0,980	0,179	1,803	0,291	1,012	0,183	1,762	0,279	0,985	0,179	1,797	0,287
27	0,909	0,220	-2,904	0,584	0,926	0,224	-2,861	0,569	0,924	0,224	-2,867	0,572	0,947	0,229	-2,812	0,553	0,934	0,226	-2,843	0,566
28	0,938	0,284	-3,241	0,777	0,962	0,289	-3,177	0,747	0,965	0,290	-3,170	0,746	0,986	0,297	-3,117	0,726	0,998	0,302	-3,088	0,719
29	0,906	0,217	-2,396	0,475	0,922	0,219	-2,363	0,461	0,921	0,218	-2,367	0,460	0,923	0,218	-2,362	0,458	0,924	0,217	-2,360	0,455
30	0,752	0,247	-4,490	1,329	0,750	0,245	-4,500	1,329	0,752	0,244	-4,486	1,315	0,751	0,245	-4,493	1,327	0,756	0,244	-4,468	1,307
31	0,792	0,171	-2,208	0,436	0,815	0,174	-2,158	0,418	0,817	0,175	-2,153	0,416	0,823	0,176	-2,140	0,412	0,812	0,174	-2,164	0,419
32	0,490	0,125	-3,356	0,851	0,497	0,125	-3,313	0,831	0,496	0,125	-3,321	0,832	0,493	0,124	-3,338	0,837	0,492	0,124	-3,346	0,838
33	0,595	0,126	0,819	0,250	0,594	0,126	0,821	0,251	0,602	0,127	0,812	0,247	0,609	0,127	0,804	0,244	0,614	0,127	0,797	0,241
34	1,043	0,182	-0,551	0,130	1,054	0,183	-0,548	0,128	1,058	0,183	-0,546	0,127	1,052	0,182	-0,549	0,128	1,050	0,181	-0,550	0,129
35	0,860	0,179	-1,036	0,215	0,845	0,176	-1,051	0,219	0,835	0,175	-1,060	0,222	0,818	0,172	-1,077	0,227	0,820	0,171	-1,076	0,226
36	0,349	0,098	5,952	1,726	0,346	0,097	5,996	1,744	-	-	-	-	-	-	-	-	-	-	-	-
37	0,382	0,096	-3,224	0,858	0,375	0,094	-3,281	0,877	0,374	0,094	-3,285	0,877	-	-	-	-	-	-	-	-
38	1,073	0,243	-2,881	0,510	1,070	0,243	-2,885	0,514	1,075	0,245	-2,876	0,513	1,047	0,240	-2,932	0,535	1,056	0,240	-2,913	0,525
39	1,337	0,282	-1,580	0,239	1,327	0,280	-1,588	0,241	1,339	0,282	-1,580	0,237	1,322	0,277	-1,593	0,240	1,300	0,274	-1,610	0,245
40	1,530	0,324	-1,742	0,244	1,526	0,321	-1,745	0,243	1,533	0,320	-1,741	0,241	1,532	0,319	-1,742	0,241	1,542	0,324	-1,737	0,240
41	1,178	0,279	-2,186	0,373	1,194	0,283	-2,166	0,367	1,213	0,285	-2,143	0,358	1,202	0,283	-2,156	0,362	1,199	0,282	-2,161	0,365
42	1,350	0,225	-1,239	0,167	1,363	0,227	-1,233	0,165	1,367	0,226	-1,231	0,164	1,369	0,226	-1,230	0,164	1,371	0,225	-1,231	0,162
43	1,056	0,200	-1,743	0,279	1,065	0,201	-1,733	0,275	1,072	0,202	-1,725	0,273	1,065	0,201	-1,734	0,275	1,058	0,198	-1,743	0,276
44	0,716	0,178	4,285	0,992	0,712	0,177	4,306	0,999	0,702	0,175	4,360	1,021	0,702	0,174	4,361	1,016	0,702	0,173	4,358	1,013

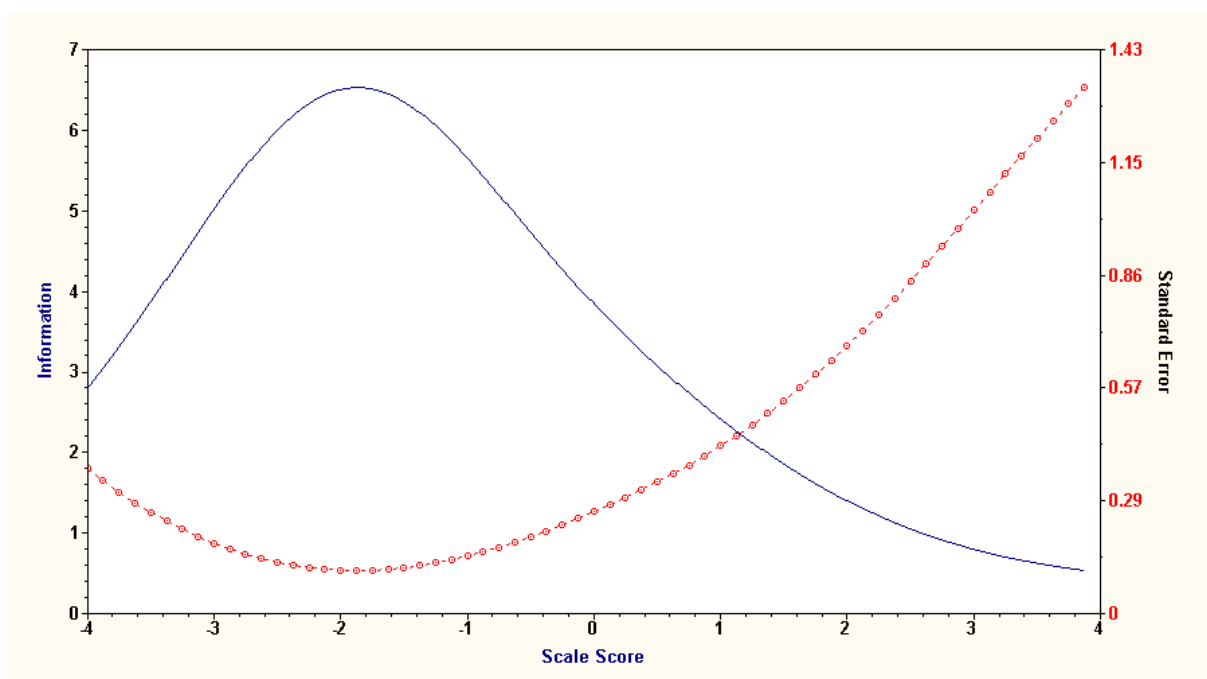
Item	39 itens				38 itens				37 itens				36 itens				35 itens			
	Par. a	erro	Par. b	erro	Par. a	erro	Par. b	erro	Par. a	erro	Par. b	erro	Par. a	erro	Par. b	erro	Par. a	erro	Par. b	erro
01	0.588	0.154	-3.660	0.918	0.562	0.148	-3.812	0.973	0.536	0.143	-3.978	1.035	0.527	0.141	-4.039	1.060	0.525	0.141	-4.058	1.066
02	0.773	0.153	-1.520	0.299	0.775	0.153	-1.516	0.298	0.762	0.149	-1.537	0.301	0.765	0.150	-1.534	0.300	0.753	0.147	-1.553	0.306
03	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
04	0.980	0.172	-1.355	0.226	0.962	0.171	-1.374	0.232	0.938	0.167	-1.400	0.237	0.923	0.165	-1.418	0.242	0.931	0.166	-1.409	0.239
05	0.477	0.110	0.106	0.228	0.462	0.107	0.109	0.235	-	-	-	-	-	-	-	-	-	-	-	-
06	0.816	0.192	-2.957	0.608	0.819	0.192	-2.949	0.605	0.812	0.191	-2.969	0.614	0.807	0.189	-2.984	0.617	0.813	0.190	-2.966	0.611
07	0.635	0.125	1.831	0.380	0.643	0.126	1.812	0.374	0.620	0.124	1.868	0.393	0.617	0.123	1.876	0.396	0.615	0.123	1.884	0.399
08	0.494	0.109	-0.565	0.246	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
09	0.504	0.112	-0.745	0.267	0.456	0.104	-0.814	0.298	0.434	0.101	-0.852	0.315	-	-	-	-	-	-	-	-
10	0.850	0.161	-0.851	0.189	0.819	0.157	-0.876	0.197	0.804	0.154	-0.889	0.201	0.790	0.152	-0.901	0.204	0.816	0.156	-0.879	0.197
11	0.855	0.155	-1.481	0.269	0.849	0.154	-1.489	0.272	0.840	0.153	-1.503	0.275	0.826	0.152	-1.522	0.280	0.818	0.150	-1.534	0.284
12	0.777	0.209	4.716	1.163	0.774	0.210	4.731	1.175	0.772	0.211	4.740	1.189	0.765	0.209	4.775	1.201	0.787	0.213	4.661	1.152
13	0.559	0.119	0.346	0.212	0.550	0.117	0.351	0.215	0.534	0.115	0.361	0.221	0.535	0.115	0.360	0.221	-	-	-	-
14	0.983	0.228	-2.397	0.443	1.003	0.232	-2.361	0.430	1.023	0.237	-2.327	0.418	1.040	0.242	-2.301	0.411	1.046	0.241	-2.292	0.403
15	0.904	0.231	-2.998	0.631	0.926	0.234	-2.939	0.606	0.924	0.231	-2.945	0.604	0.939	0.233	-2.906	0.588	0.947	0.234	-2.889	0.579
16	1.062	0.177	0.159	0.116	1.003	0.168	0.167	0.121	0.987	0.167	0.170	0.123	0.972	0.164	0.172	0.124	0.944	0.161	0.178	0.127
17	1.430	0.363	-2.486	0.389	1.461	0.378	-2.452	0.382	1.475	0.389	-2.439	0.383	1.477	0.396	-2.435	0.388	1.452	0.394	-2.463	0.401
18	0.770	0.186	-2.660	0.575	0.770	0.184	-2.662	0.573	0.778	0.185	-2.641	0.564	0.780	0.185	-2.634	0.560	0.793	0.186	-2.599	0.544
19	1.282	0.238	0.446	0.116	1.304	0.242	0.442	0.114	1.340	0.246	0.436	0.111	1.369	0.249	0.430	0.109	1.378	0.249	0.431	0.109
20	0.764	0.154	-0.287	0.164	0.780	0.156	-0.281	0.160	0.764	0.154	-0.284	0.164	0.754	0.153	-0.286	0.166	0.739	0.150	-0.289	0.169
21	1.188	0.297	-2.521	0.443	1.204	0.304	-2.497	0.440	1.209	0.304	-2.491	0.437	1.223	0.311	-2.471	0.435	1.214	0.313	-2.485	0.444
22	1.047	0.182	0.161	0.118	1.026	0.179	0.163	0.121	0.992	0.174	0.169	0.124	0.969	0.171	0.172	0.127	0.962	0.169	0.175	0.128
23	1.026	0.224	-2.104	0.363	1.065	0.230	-2.049	0.343	1.104	0.238	-1.999	0.325	1.137	0.246	-1.958	0.314	1.141	0.243	-1.955	0.309
24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	0.953	0.168	-0.964	0.180	0.959	0.168	-0.960	0.178	0.979	0.170	-0.947	0.173	0.989	0.172	-0.941	0.171	0.994	0.172	-0.938	0.170
26	0.994	0.180	1.785	0.284	1.022	0.182	1.750	0.274	1.042	0.186	1.728	0.269	1.067	0.189	1.700	0.261	1.028	0.184	1.746	0.274
27	0.901	0.219	-2.925	0.597	0.904	0.220	-2.915	0.594	0.909	0.222	-2.903	0.591	0.911	0.222	-2.899	0.590	0.910	0.223	-2.902	0.591
28	0.979	0.299	-3.133	0.742	1.007	0.304	-3.067	0.711	1.032	0.314	-3.012	0.692	1.067	0.320	-2.936	0.654	1.082	0.322	-2.908	0.639
29	0.918	0.216	-2.373	0.461	0.936	0.220	-2.337	0.448	0.972	0.226	-2.272	0.422	0.999	0.229	-2.228	0.406	1.020	0.232	-2.195	0.392
30	0.760	0.246	-4.446	1.300	0.758	0.243	-4.456	1.292	0.758	0.239	-4.456	1.273	0.752	0.234	-4.487	1.277	0.737	0.227	-4.565	1.295
31	0.807	0.174	-2.175	0.423	0.831	0.179	-2.125	0.408	0.855	0.182	-2.079	0.391	0.881	0.188	-2.031	0.377	0.879	0.186	-2.035	0.376
32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
33	0.644	0.131	0.766	0.228	0.638	0.130	0.772	0.230	0.639	0.129	0.772	0.231	0.635	0.128	0.776	0.232	0.622	0.126	0.791	0.237
34	1.033	0.179	-0.555	0.131	1.062	0.181	-0.546	0.127	1.074	0.180	-0.542	0.126	1.077	0.179	-0.542	0.125	1.072	0.176	-0.543	0.125
35	0.831	0.172	-1.065	0.222	0.843	0.173	-1.053	0.217	0.823	0.169	-1.073	0.224	0.808	0.167	-1.088	0.228	0.796	0.164	-1.101	0.232
36	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
37	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
38	1.061	0.239	-2.904	0.521	1.070	0.241	-2.884	0.515	1.077	0.240	-2.871	0.506	1.060	0.235	-2.905	0.517	1.056	0.236	-2.915	0.521
39	1.292	0.270	-1.617	0.246	1.287	0.269	-1.621	0.246	1.289	0.270	-1.622	0.245	1.287	0.271	-1.624	0.246	1.283	0.268	-1.629	0.245
40	1.537	0.323	-1.740	0.240	1.547	0.321	-1.735	0.237	1.539	0.320	-1.742	0.240	1.551	0.325	-1.735	0.239	1.530	0.311	-1.750	0.238
41	1.185	0.280	-2.179	0.372	1.209	0.287	-2.149	0.363	1.225	0.290	-2.131	0.356	1.249	0.296	-2.104	0.348	1.279	0.301	-2.073	0.334
42	1.364	0.224	-1.234	0.163	1.376	0.224	-1.229	0.160	1.395	0.226	-1.221	0.157	1.409	0.227	-1.215	0.155	1.394	0.223	-1.224	0.156
43	1.065	0.199	-1.735	0.273	1.088	0.202	-1.710	0.264	1.095	0.203	-1.703	0.262	1.100	0.204	-1.698	0.260	1.117	0.207	-1.681	0.255
44	0.694	0.173	4.401	1.033	0.695	0.173	4.394	1.032	0.691	0.173	4.419	1.047	0.688	0.173	4.434	1.051	0.679	0.172	4.483	1.074

Item	34 itens				33 itens				32 itens			
	Par. a	erro	Par. b	erro	Par. a	erro	Par. b	erro	Par. a	erro	Par. b	erro
01	0.521	0.138	-4.086	1.066	-	-	-	-	-	-	-	-
02	0.748	0.148	-1.561	0.309	0.756	0.149	-1.549	0.305	0.760	0.149	-1.541	0.302
03	-	-	-	-	-	-	-	-	-	-	-	-
04	0.896	0.161	-1.450	0.249	0.873	0.159	-1.480	0.259	0.885	0.160	-1.463	0.254
05	-	-	-	-	-	-	-	-	-	-	-	-
06	0.825	0.193	-2.933	0.601	0.826	0.195	-2.931	0.602	0.824	0.194	-2.939	0.605
07	0.646	0.126	1.806	0.372	0.645	0.126	1.808	0.373	-	-	-	-
08	-	-	-	-	-	-	-	-	-	-	-	-
09	-	-	-	-	-	-	-	-	-	-	-	-
10	0.800	0.152	-0.892	0.201	0.780	0.148	-0.910	0.206	0.787	0.148	-0.902	0.204
11	0.818	0.151	-1.534	0.284	0.819	0.151	-1.533	0.283	0.798	0.148	-1.565	0.293
12	0.789	0.214	4.647	1.150	0.787	0.215	4.657	1.162	0.781	0.217	4.685	1.195
13	-	-	-	-	-	-	-	-	-	-	-	-
14	1.080	0.246	-2.241	0.382	1.080	0.247	-2.241	0.383	1.044	0.238	-2.298	0.400
15	0.950	0.234	-2.881	0.576	0.946	0.231	-2.890	0.575	0.934	0.225	-2.920	0.579
16	0.957	0.163	0.176	0.126	0.942	0.161	0.179	0.128	0.967	0.164	0.177	0.125
17	1.445	0.403	-2.472	0.412	1.442	0.408	-2.476	0.419	1.424	0.402	-2.500	0.427
18	0.784	0.183	-2.624	0.551	0.789	0.184	-2.611	0.545	0.768	0.178	-2.672	0.561
19	1.242	0.225	0.461	0.118	1.239	0.225	0.462	0.118	1.293	0.232	0.452	0.113
20	0.782	0.155	-0.274	0.160	0.777	0.154	-0.275	0.161	0.750	0.152	-0.282	0.167
21	1.193	0.310	-2.517	0.460	1.191	0.312	-2.520	0.465	1.176	0.305	-2.544	0.467
22	0.919	0.164	0.182	0.133	0.920	0.164	0.183	0.133	0.914	0.162	0.185	0.133
23	1.162	0.251	-1.933	0.304	1.180	0.254	-1.914	0.297	1.152	0.246	-1.946	0.304
24	-	-	-	-	-	-	-	-	-	-	-	-
25	0.959	0.166	-0.962	0.177	0.971	0.167	-0.953	0.174	0.975	0.17	-0.949	0.173
26	1.037	0.187	1.736	0.273	1.067	0.189	1.702	0.261	1.086	0.193	1.682	0.257
27	0.951	0.230	-2.807	0.552	0.960	0.233	-2.785	0.546	0.935	0.225	-2.845	0.562
28	1.094	0.321	-2.887	0.623	1.111	0.326	-2.854	0.612	1.106	0.323	-2.866	0.610
29	1.037	0.236	-2.169	0.383	1.051	0.236	-2.150	0.373	1.052	0.234	-2.149	0.368
30	0.754	0.237	-4.477	1.282	0.755	0.235	-4.473	1.272	0.752	0.231	-4.490	1.260
31	0.868	0.182	-2.055	0.378	0.887	0.185	-2.022	0.367	0.872	0.181	-2.049	0.375
32	-	-	-	-	-	-	-	-	-	-	-	-
33	-	-	-	-	-	-	-	-	-	-	-	-
34	1.062	0.173	-0.546	0.126	1.072	0.173	-0.543	0.125	1.090	0.174	-0.535	0.123
35	0.781	0.162	-1.117	0.236	0.785	0.162	-1.113	0.234	0.757	0.157	-1.145	0.245
36	-	-	-	-	-	-	-	-	-	-	-	-
37	-	-	-	-	-	-	-	-	-	-	-	-
38	1.051	0.231	-2.927	0.519	1.049	0.233	-2.930	0.523	1.036	0.228	-2.959	0.528
39	1.301	0.271	-1.616	0.239	1.285	0.266	-1.629	0.241	1.276	0.265	-1.637	0.244
40	1.556	0.313	-1.736	0.231	1.567	0.316	-1.730	0.229	1.544	0.311	-1.746	0.233
41	1.310	0.304	-2.044	0.321	1.322	0.307	-2.032	0.317	1.277	0.293	-2.015	0.329
42	1.463	0.234	-1.193	0.148	1.453	0.233	-1.198	0.148	1.437	0.227	-1.205	0.150
43	1.115	0.206	-1.685	0.256	1.095	0.202	-1.707	0.261	1.107	0.201	-1.694	0.255
44	0.694	0.175	4.401	1.046	0.697	0.174	4.382	1.033	0.690	0.175	4.420	1.061

APÊNDICE F – Curva de Informação dos conjuntos de itens



Curva de informação total dos 44 item iniciais.



Curva de informação total dos 32 item finais.