

Todos os usuários afirmaram que, enquanto tentavam selecionar a opção em alguma *drop-down-list*, o sistema fazia um *refresh* (atualizar) e a *drop-down-list* se retraiía novamente, o que os obrigava a clicar e procurar o item a ser selecionado.

Alguns usuários disseram que as mensagens poderiam ser mais bem redigidas. Como por exemplo: “Atenção, as sessões abaixo vão ocorrer após hoje, 27 de agosto”.

Apesar dos problemas encontrados nessa avaliação, os usuários conseguiram realizar todas as tarefas. Pode-se perceber, portanto, que as observações e os comentários feitos durante a realização das tarefas valeram mais do que um registro da interação entre o usuário e o sistema que verificasse o tempo gasto em cada tarefa e sua completude. Nesse caso, vimos que avaliação cooperativa possibilitou a detecção de uma série de problemas de usabilidade, de diferentes graus de gravidade, a baixo custo, em pouco tempo e mobilizando poucos profissionais e usuários para sua realização.

Cabe ressaltar os problemas detectados durante a navegação para a busca de informações. Outras etapas específicas da compra *online* não foram consideradas. Percebe-se assim a dificuldade para encontrar e realmente efetuar a compra no *site* seria ainda maior, pois deveríamos considerar os aspectos da interação entre usuário-formulário, cadastro do usuário, preenchimento de informações pessoais e bancárias, entre outras.

A cada problema de interação, que isolado poderia não apresentar gravidade, ocorre uma crescente insatisfação do usuário. Em certos casos, isso pode acarretar a desistência da compra e, em outros até a rejeição do *site*.

9. Teste de usabilidade

Segundo Travis (2003), os testes de usabilidade têm suas raízes na psicologia experimental, na qual se pede aos participantes que executem uma ou mais tarefas bem determinadas, e se faz uma análise estatística dos resultados. Para a maioria das organizações esse esquema não é viável, portanto, ao longo dos últimos dez ou 15 anos surgiram diferentes variações.

O teste de usabilidade é um método, empregado na ergonomia e na interação humano-computador, para testar e avaliar a usabilidade de produtos e sistemas, a partir da observação dos usuários durante a interação.

O foco do testes de usabilidade está no comportamento observável – o que os usuários fazem mais até do que o que dizem que fazem.

De acordo com Rodrigues (2001), foram os testes de usabilidade os responsáveis por dar à *web* a credibilidade que lhe faltava, colocando fim ao “achismo” dos primeiros anos.

Segundo Dumas e Redish (1999), os testes de usabilidade se diferenciam dos beta-testes, pois, apesar de um beta-teste típico ser uma liberação precoce de um produto para uns poucos usuários e de ter pessoas reais usando o produto nos ambientes reais, executando tarefas reais, o beta-teste raramente propicia qualquer

informação útil sobre usabilidade. A maioria das companhias tem achado que o beta-teste é muito pequeno, muito assistemático e muito tardio para ser o principal teste de usabilidade.

9.1 Objetivos dos testes de usabilidade

O objetivo principal de um teste de usabilidade é melhorar a facilidade de uso de um produto. Utilizando testes de usabilidade, a equipe de desenvolvimento pode saber imediatamente se as pessoas compreendem o design como se suporia que entendessem.

De acordo com Marcus (2002), a observação de poucos usuários lidando com o produto tem um impacto muito maior sobre as atitudes dos desenvolvedores do que muitas horas de discussão sobre a importância da usabilidade ou sobre como compreender o usuário.

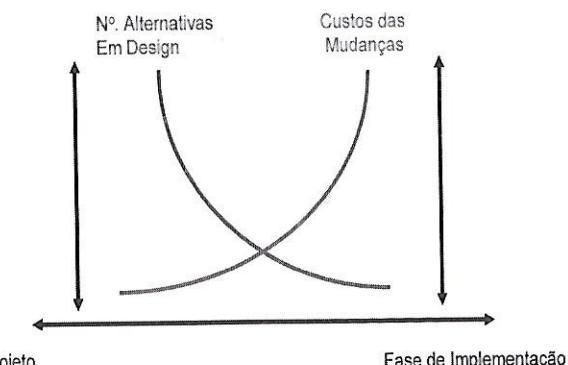
Dumas e Redish (1999) acrescentam, como benefícios dos testes de usabilidade, a mudança de atitudes das pessoas sobre os usuários e a mudança no processo de design e desenvolvimento.

De acordo com Kuniavsky (2003), embora o teste possa certamente gerar novas idéias para uma próxima revisão do produto, a técnica é melhor utilizada durante o ciclo de desenvolvimento iterativo, tanto para testar a usabilidade de aspectos específicos quanto para investigar novas idéias.

Quanto mais cedo os testes forem introduzidos no processo de desenvolvimento, menores serão os custos e melhores os resultados. É absolutamente viável realizar testes de usabilidade em protótipos de todo o tipo, bem como outras técnicas que trazem à luz a visão do usuário sobre a interação (Fernandez, 2005).

Marcus (2002) apresenta em seu artigo *Return on Investment for Usable User-Interface Design: Examples and Statistics* o diagrama a seguir, que ilustra as curvas relativas ao custo das mudanças

e alterações *versus* o número de alternativas de design possíveis, em função das fases de pesquisas pré-liminares, desenvolvimento e utilização. Verifica-se no diagrama que, nas fases iniciais do projeto, existem muitas alternativas de design a baixos custos. À medida que o sistema é desenvolvido, o número de alternativas de design diminui e seus custos se tornam mais elevados.



Esquema 13 – Alternativas de design x fases de desenvolvimento x custos de mudanças

Para Dumas e Redish (1999), o teste de usabilidade é apropriado interativamente desde o pré-design (testar um produto similar ou uma versão mais antiga), por meio do design inicial (testar protótipos) e pelo desenvolvimento (testar aspectos diferentes e retestar as alterações).

A partir dos testes de usabilidade, avalia-se o desempenho dos usuários típicos na realização de tarefas cuidadosamente preparadas, por sua vez típicas daqueles para os quais o sistema foi projetado. O desempenho dos usuários é geralmente avaliado no que diz respeito ao número de erros e ao tempo para completar a tarefa (Preece, Rogers e Sharp, 2005).

Embora possam ocorrer amplas variações sobre como e onde um teste de usabilidade é conduzido, todos os testes compartilham cinco características:

1. Objetivo principal.
2. Os participantes representam usuários reais.
3. Os participantes executam tarefas reais.
4. Observa-se e regista-se o que os participantes fazem e dizem.
5. Analisam-se os dados, diagnosticam-se os problemas reais e, então, recomendam-se alterações para consertar tais problemas.

9.2 Planejamento do teste de usabilidade

De acordo com Badre (2002), o processo de condução do teste segue as seguintes etapas:

1. **Planejamento do teste** (determinar o objetivo do teste e as tarefas).
2. **Organização dos materiais** (*site, script* do teste, formulários de consentimento, e outros materiais).
3. **Preparação do local** (local, equipamento e sala).
4. **Teste piloto** (com um ou dois colaboradores).
5. **Recrutamento os usuários** (selecionar e agendar as horas de teste).
6. **Condução do teste.**
7. **Análise dos resultados** (revisar os problemas encontrados, priorizar os problemas baseados na freqüência e na severidade, e identificar as soluções possíveis).
8. **Correção do site** (fazer as alterações modificadas e testar mais uma vez).

Ao planejar um teste de usabilidade na *web*, há certos aspectos que devem ser avaliados com maior ênfase. Um plano para o teste de usabilidade deve incluir o propósito do *website*, perfis do

público-alvo e cenários típicos do uso. Além disso, outros aspectos devem ser considerados, tais como:

- Tipos de usuários para o teste.
- Número de usuários para o teste.
- Localização do teste.
- Tarefas a serem executadas no teste.
- Simulação das condições de uso.

De acordo com Preece, Rogers e Sharp (2005), geralmente os testes são realizados em condições controladas, como em laboratório. Não é permitida a entrada de visitantes e as chamadas telefônicas são canceladas, não havendo a possibilidade de se falar com colegas, verificar mensagens ou realizar qualquer uma das outras tarefas que a maioria de nós realiza corriqueiramente na vida normal. Todas as ações do participante são registradas – toques nas teclas, comentários, pausas, expressões etc. – de modo a serem usadas como dados. Apesar de o ambiente do laboratório de usabilidade ser controlado, os usuários geralmente esquecem que estão sendo observados.

Kuniavsky (2003) apresenta a seguinte programação para o planejamento de testes de usabilidade:

- 2 semanas – Determinar o público-alvo e começar o recrutamento.
- + 2 semanas – Determinar o conjunto de aspectos a serem testados.
- + 1 semana – Fazer a primeira versão do *script*, construir as tarefas-teste, discutir com o grupo de desenvolvimento e conferir o recrutamento.
- + 3 dias – Escrever uma segunda versão do *script*, revisar as tarefas, discutir com o grupo e completar o recrutamento.
- + 2 dias – Completar o guia (*script*), agendar os testes práticos e conferir todo o equipamento.

- + 1 dia – Fazer o teste prático e ajustar o guia e as tarefas conforme a necessidade.
- + 1 dia – Discutir com os observadores e coletar cópias e anotações.
- + 3 dias – Olhar as fitas e fazer anotações.
- + 1 semana – Combinar as anotações e escrever a análise.
- + 1 semana – Apresentar ao grupo de desenvolvimento, discutir e anotar as diretrizes para a próxima pesquisa.

9.3 Variações dos testes de usabilidade

Na maioria dos testes de usabilidade, observa-se um participante de cada vez trabalhando com o produto. Geralmente, o pesquisador deixa o participante sozinho e observa do canto da sala ou por trás de um espelho unidirecional. Ele intervém somente quando a pessoa pede auxílio. Cada intervenção é anotada como um pedido de assistência para análises posteriores.

Os testes são estruturados dessa maneira para simular o que acontece quando os usuários individuais utilizam os produtos nos seus escritórios ou em suas casas. Entretanto, algumas vezes pode-se querer mudar essas técnicas. Duas idéias que muitos grupos acham úteis são:

- a) Co-descoberta – Tendo dois participantes trabalhando juntos.
- b) Intervenção ativa – Exercendo um papel mais ativo no teste.

Segundo Preece, Rogers e Sharp (2005), em alguns casos é pertinente solicitar aos participantes que exteriorizem seus processos de raciocínio. Essa técnica foi desenvolvida por Erikson e Simon, em 1985, e consiste em exigir que as pessoas falem alto tudo o que estiverem pensando.

9.4 Laboratório de usabilidade

De acordo com Rubin (1994), para muitos dos envolvidos na implementação de um programa de avaliação da usabilidade, a disciplina se tornou sinônimo de um laboratório muito potente, bem equipado e caro. Para muitas organizações, o laboratório de usabilidade (no sentido da estrutura física) tornou-se mais proeminente e mais importante que o próprio processo de avaliação. Com o intuito de impressionar clientes e competidores com relação ao seu comprometimento com a usabilidade, algumas dessas organizações criaram “palácios amedrontadores de bruxaria *high-tech*” em vez de estabelecer os fundamentos para um programa de avaliação.

De acordo com Badre (2002), antigamente usava-se um laboratório formal de testes. Atualmente, devido aos custos, realiza-se o teste em um ambiente mais informal. Isso não apenas é mais barato, mas também permite ao avaliador perceber qual o tipo de problema que os usuários encontram no seu contexto habitual.

Rubin (1994) ressalta que, independentemente de estar iniciando um grande programa de avaliação ou simplesmente testando o seu próprio produto, não é necessário um laboratório caro e sofisticado para atingir seus objetivos. Na verdade, começar pequeno direciona o foco em todos os outros fatores que tornam um programa bem-sucedido.

De acordo com Preece, Rogers e Sharp (2005), seja em um laboratório real ou montado, um dos problemas da observação está no fato de o observador não saber o que os usuários estão pensando, podendo inferir somente a partir do que vê.

Segundo Rubin (1994), alguns ambientes de laboratórios de usabilidade são simplórios, outros mais sofisticados – com siste-

mas de monitoramento, vidros unidirecionais e equipamentos de áudio e vídeo para comunicação entre os participantes, observadores e pesquisadores.

A seguir são apresentados dois esquemas de laboratórios de usabilidade. O esquema simples tem apenas uma sala, com um microcomputador, e câmera de vídeo. O esquema clássico apresenta duas salas. Uma para o participante e uma para o monitor, equipe de teste e observadores. A visualização do teste é feita através de uma janela com espelho unidirecional. A equipe de teste monitora a realização das tarefas por meio de monitores com imagens do computador e captadas por duas câmeras de vídeo instaladas no laboratório.

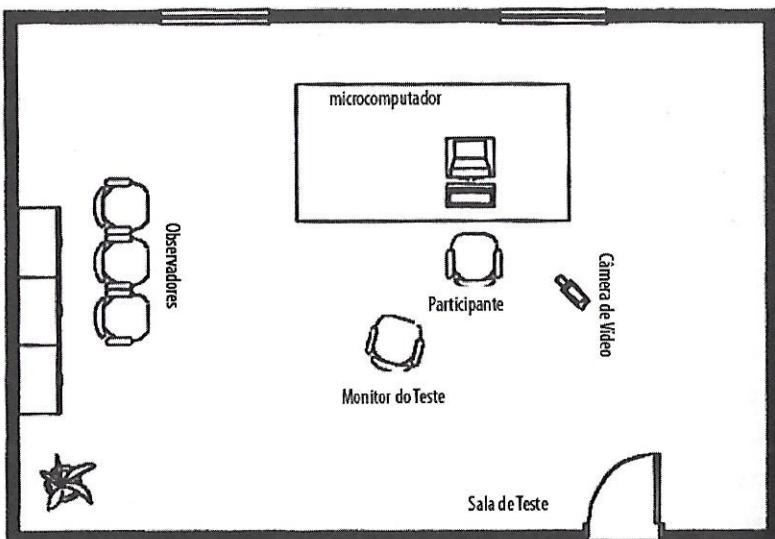


Figura 20 – Laboratório de usabilidade:
Esquema simples (sala única)

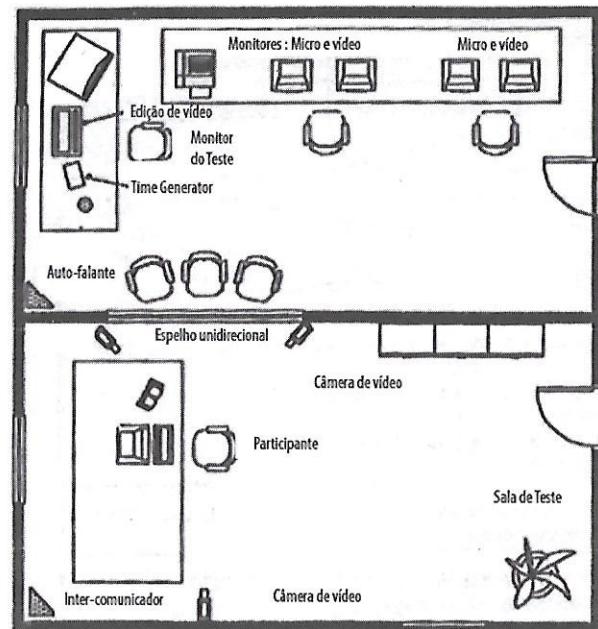


Figura 21 – Laboratório de usabilidade: Esquema clássico

9.4.1 Equipamentos para os testes de usabilidade

Para Travis (2003), o equipamento do laboratório de usabilidade é um híbrido entre os laboratórios de psicologia experimental e o laboratório de um grupo de foco. Assim, por exemplo, o participante do teste deve ficar sentado no estúdio e ser gravado por vídeo e câmeras enquanto trabalha no *website*. O administrador e qualquer observador interessado devem se sentar na sala de observação, de onde possam ver o usuário no vídeo ou através de um espelho unidirecional.

Além do equipamento de vídeo, serão necessárias ferramentas para coleta de dados. Essas podem variar em sofisticação, desde softwares de *logs* (para medir o tempo das tarefas e sub-ta-

refas), até gravadores do movimento dos olhos, para determinar para onde, precisamente, o participante está olhando. Contudo, para a maioria dos testes de usabilidade, é preciso apenas de folhas de resposta cuidadosamente preparadas, lápis e um cronômetro (Travis, 2003).

Para Dumas e Redish (1999), um teste de usabilidade pode dispensar alguns recursos usualmente citados pela literatura, tais como: laboratório com espelho unidirecional, software de captura de tela, videocassete e formulário de desempenho. Segundo o autor, cada um desses itens é útil, mas não é necessário para que um teste de usabilidade seja bem sucedido.

Krug (2006) também ressalta que um laboratório de usabilidade ultramoderno nem sempre é preciso e que sua ausência não pode servir como desculpa para evitar a realização dos testes de usabilidade. Segundo o pesquisador, na maioria das vezes a gravação de tudo o que acontece na tela por meio de um software como o *camtasia*, fornece informações valiosas além de permitir o exame e o compartilhamento entre membros da equipe de forma mais simples e ágil.

9.5 Recrutamento dos participantes

Rubin (1994) afirma que a escolha do número de participantes selecionados para os testes depende de muitos fatores, incluindo:

- O grau de confiança necessário nos resultados.
- O número de recursos disponíveis para instituir e conduzir o teste.
- A disponibilidade do tipo de participante requerido.
- A duração de cada sessão do teste.
- O tempo necessário para preparar o teste.

De acordo com Kuniavsky (2003), as melhores pessoas para convidar são aquelas que precisarão, num futuro próximo, do

serviço disponibilizado, ou aquelas que, num passado recente, utilizaram um produto concorrente.

Testar a usabilidade é ver o projeto através dos olhos dos participantes. Por isso eles têm que ser pertencentes ao público-alvo do sistema ou produto. No caso de participantes com experiência acima do público-alvo, podem não vivenciar os problemas que o usuário real encontraria. Se eles não têm experiência suficiente, podem se atrapalhar com coisas que os usuários resolveriam facilmente. Um erro comum é focalizar-se na demografia. (tal como idade e renda) e não atentar para as distinções que levam os usuários a se comportar diferentemente, tais como, a sua fluência no conteúdo do projeto. O risco é perder problemas críticos que seriam facilmente consertados justamente porque os participantes recrutados não vieram a encontrá-los.

Para Rubin (1994), se for utilizado um esquema matricial, deve-se testar ao menos quatro participantes por grupo de tratamento. Num exemplo sobre acesso a sistemas bancários, se há suspeita de que 75% dos usuários eventuais sejam novatos, deve-se certificar que a maioria dos participantes pertença a essa categoria.

Segundo Dumas e Redish (1999), um teste de usabilidade deve ter no mínimo duas ou três pessoas representando um subgrupo de pessoas, de modo a evitar uma análise de comportamento idiossincrático.

Nielsen (1993) sugere que sejam utilizados cinco participantes em testes de usabilidade, pois apresenta melhor relação custo x benefício.

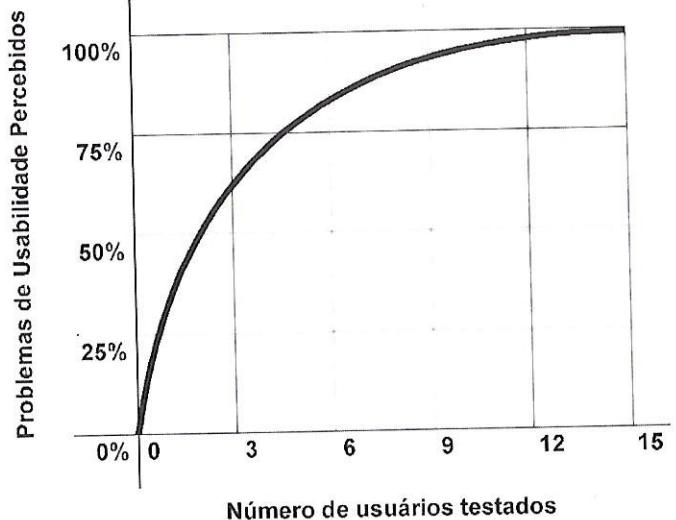


Figura 22 – Número de usuários testados x problemas de usabilidade percebidos

Com cinco participantes, 85% dos problemas são detectados, enquanto que, segundo Virzi (1989) apud Nielsen (1993), para descobrir 90% dos problemas, seria necessário o dobro de participantes.

Krug (2000) faz uma crítica irreverente, desmistificando um pouco a necessidade da determinação de um número exato de participantes, afirmando que testar um usuário é 100% melhor do que não testar nenhum.

Embora existam algumas divergências entre o número exato de participantes, verifica-se que a opinião dos pesquisadores de usabilidade citados nesse referencial teórico transita entre três a dez usuários. A escolha do número de usuários deve ser feita entre essa faixa, e levar em consideração, além das justificativas apresentadas

pelos supracitados autores, aspectos técnicos, restrições financeiras e de tempo, e os objetivos do teste de usabilidade.

A seguir é apresentado um quadro comparativo entre o número de usuários propostos por diversos pesquisadores de HCI (Santa-Rosa, 2005):

	Número de pessoas	Justificativa
Nielsen	5	Com cinco participantes, 85% dos problemas de usabilidade são descobertos.
Rubin	8	Embora possa ser evidenciada, a maioria dos problemas com quatro participantes poderá ignorar alguns erros sérios de usabilidade.
Dumas	de 6 a 12	Três participantes é o mínimo para um subgrupo.
Krug	3 ou 4	Com três ou quatro, a maioria dos problemas é detectada, tornando possível fazer a análise no mesmo dia. Testar um participante é 100% melhor do que testar nenhum.
Virzi	10	Com dez participantes, 90% dos problemas são descobertos.

http://www.multipolo.com.br/usabilidade/participantes_usab.htm

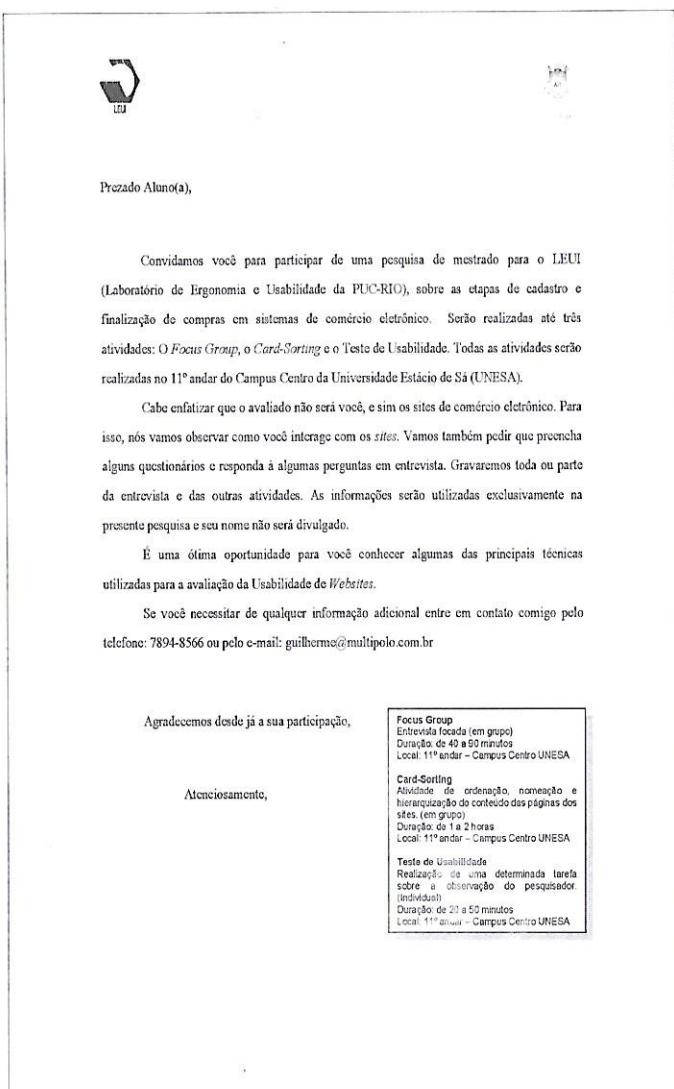
Quadro 4 – Quadro comparativo entre número de usuários por subgrupo para a realização de testes de usabilidade

Sugere-se enviar uma carta-convite aos participantes, na qual fiquem claros os objetivos do teste e sejam apresentadas informações sobre método utilizado, data, local e duração aproximada de cada teste.

De acordo com Rubin (1994), a carta-convite para as atividades da avaliação de usabilidade deve incluir:

- O tipo de produto (para que o usuário possa perceber o quanto ele será importante para a realização do teste).
- O ambiente (informando que é um laboratório com câmeras de vídeo e que poderá lhe ser solicitado “pensar alto” enquanto interage com o produto).
- Quanto tempo aproximadamente vai durar o teste.

A seguir é apresentada a carta-convite redigida para a realização de uma pesquisa sobre formulários eletrônicos que utilizou três técnicas: *grupo de foco*, *cardsorting* e teste de usabilidade (Santa-Rosa, 2005; Santa-Rosa e Moraes, 2006b).



9.6 Especificação das tarefas e cenários

As tarefas que o pesquisador estipulará para que o usuário execute no teste devem ser aquelas que ele executará com o produto em seu emprego ou em sua casa. Isso significa que o pesquisador deve compreender o emprego do usuário e as tarefas para as quais aquele produto seja relevante.

Além de serem realísticas e relevantes para o usuário, as tarefas que serão incluídas no teste devem estar relacionadas com os objetivos e expectativas da pesquisa, e ter grande probabilidade de revelar um problema de usabilidade.

As boas tarefas devem ser:

1. Razoáveis – Devem ser típicas das coisas que as pessoas fazem.
2. Descritas em termos de objetivos terminais.
3. Específicas – Descrever especificamente a tarefa.
4. Factíveis – Apresentadas numa seqüência realista.
5. Domínio neutro – Seleção de conteúdo corriqueiro.
6. Tamanho razoável – De modo que possa ser executada no tempo previsto.

De acordo com Kuniavsky (2003), podem-se estabelecer tarefas destinadas a testes de usabilidade competitivos. Embora comparar duas interfaces gaste mais tempo do que uma interface só, isso pode revelar pontos positivos e negativos em ambos os produtos. Realizar a mesma tarefa em uma interface existente e num protótipo, por exemplo, pode revelar se o novo projeto é mais funcional. Da mesma maneira, realizar a mesma tarefa ou conduzir navegações em interfaces semelhantes de dois produtos concorrentes revela as forças relativas dos dois produtos. Segundo o supracitado autor, em ambas as situações é muito importante evitar qualquer de tipo de viés na pesquisa.

Segundo Dumas e Redish (1999) devem ser criados e utilizados cenários para mostrar aos participantes o que queremos que eles façam durante o teste. Os cenários descrevem as tarefas de um modo que elimine algumas das artificialidades do teste.

A seguir apresenta-se um exemplo de cenário elaborado para um teste de usabilidade:

“Você acabou de chegar no seu escritório depois de umas curtas férias. Verifique quantos e-mails têm à sua espera. Se tiver mensagens de seus familiares, leia-as.”

O cenário, portanto, torna a tarefa mais realística. Num cenário nós damos o objetivo e qualquer informação que um usuário realmente precise ao executar a tarefa.

Um bom cenário deve ser pequeno, usar terminologia adequada ao usuário e não ser ambíguo, de modo que todos os participantes compreendam. Deve propiciar aos participantes, informações suficientes para realizar a tarefa e se relaciona diretamente às tarefas e ao contexto.

Cabe ressaltar que sempre será necessário entregar aos usuários os cenários por escrito. Um outro meio de apresentar o cenário pode ser mais realístico do que obrigá-los a lê-lo. Pode ser adequado, portanto:

- Ter membros da equipe que interpretem o papel dos usuários, supervisores ou colegas que caminhem na sala e passem as tarefas especificadas no cenário.
- Utilizar o produto para apresentar os cenários (por exemplo: testando telefones, os cenários provavelmente incluiriam chamadas telefônicas para os participantes, bem como ligações feitas por eles).
- Misturar as modalidades: sugere-se dar aos participantes cenários por escrito mas interromper com chamadas e visitas reais.

9.7 Condução dos testes

O teste de usabilidade é uma ferramenta para produzir informação. Entretanto, ele não pode, na realidade, produzir todos os tipos de informação. Os testes mais bem sucedidos são aqueles nos quais a equipe está envolvida em todas as etapas do processo. Os participantes estão atentos em cada teste, absorvendo a informação tão rapidamente quanto ela vem sem a filtragem e distorção natural que ocorre quando eles tem que ouvir os resultados de segunda ou terceira mão. Portanto é recomendado que mais um membro da equipe assista às sessões dos testes de usabilidade.

Antes do início das sessões dos testes de usabilidade é muito importante verificar todos os equipamentos, softwares e outros recursos envolvidos no teste. Deve-se verificar a conexão com a internet, sua velocidade e estabilidade assim como outras configurações que se façam necessárias.

No caso da análise de *sites* é importante limpar o cachê do navegador assim como os arquivos temporários do sistema operacional e reconfigurar todas as opções para as default para que todas estas variáveis estejam controladas.

Em alguns testes, pode-se escolher as mesmas faixas de horários, ou dias da semana, de modo a minimizar variações de desempenho da rede ou da internet.

Dependendo da sala utilizada para o teste, sugere-se um controle no horário da realização do teste, para minimizar diferenças de iluminação e reflexos.

A realização de testes-piloto permite que sejam testados cenários, compreensão das tarefas, tempo médio necessário assim como o funcionamento e a configuração dos equipamentos. Por exemplo, no teste-piloto é possível configurar o ângulo certo da

câmera de modo que filme as expressões do usuário, sem reflexos, no horário do teste.

Sugere-se que sejam elaborados *checklists* com todos os procedimentos operacionais, desde verificar se existem fitas sobressalentes até a alocar de salas ou testar o funcionamento de softwares e hardwares.

Devem-se criar planilhas para anotações das telas, dos erros, dos comentários e expressões, de modo a facilitar a tabulação e análise dos dados.

Durante o teste, sugere-se que o observador ou pesquisador tome cuidado com as palavras e expressões corporais empregadas e monitore toda a sessão com imparcialidade. Deve-se deixar os participantes vencerem suas dificuldades sozinhos e evitar ajudá-los a não ser que a tarefa tenha sido considerada abortada.

O observador ou pesquisador deve certificar-se que o participante realmente terminou uma tarefa para então passar para a seguinte.

Cabe ressaltar, que também é possível realizar testes de usabilidade com o método *thinking aloud* embora, neste caso, a duração de cada tarefa não possa ser mais levada em consideração durante a análise dos dados.

9.8 Realização de entrevistas

Preferencialmente, para cada teste que o pesquisador realiza, a sua meta deve ser compreender porque ocorreu cada erro e a dificuldade ou omissão para cada participante durante cada seção. A entrevista é a oportunidade final para atingir essa meta antes que o participante deixe o laboratório. Ela permite resolver quaisquer dúvidas que persistam depois de uma seção. Se por um lado a porção orientada à tarefa de um teste indica ou expõe o

“quê” de um desempenho, por outro lado a entrevista expõe ou indica o “porquê” do desempenho

Para Moraes (2002), a entrevista é o método interrogativo mais flexível para coleta de dados. Ela pode ser classificada como não-diretiva (sem estruturação do problema pelo investigador), focalizada (descrição livre, mas baseada em hipóteses e em certos temas), semi-estruturada (aplicada a partir de perguntas abertas) e clínica (que visa a interpretação sóciopsicológica da situação ou da personalidade dos sujeitos através e suas verbalizações).

Segundo Kuniavsky (2003), o núcleo da entrevista é uma série de tarefas a serem executadas pelo avaliador da interface (tipicamente uma pessoa que representa a audiência ideal do produto). Fitas e anotações da entrevista são mais tarde analisadas quanto ao sucesso, más interpretações, equívocos e opiniões do avaliador. Depois de executados alguns desses testes, as avaliações são comparadas e os aspectos mais comuns são listados como problemas de funcionalidade e de apresentação.

A entrevista pretende expor o processo de pensamento e raciocínio atrás de cada uma das ações dos participantes, especialmente para aquelas ações sobre as quais o observador tem dúvida. Dumas e Redish (1999) sugerem que durante a condução das entrevistas, o pesquisador ou interrogador nunca leve os participantes a se colocarem na defensiva sobre as suas ações ou suas opiniões e, enquanto entrevista um participante, o pesquisador não deve reagir às respostas dele de uma maneira ou outra.

De acordo com Kuniavsky (2003), o pesquisador deve ficar atento ao que as pessoas não fazem ou não notam do mesmo modo que observa o que elas fazem e notam.

Dumas e Redish (1999) apresenta algumas diretrizes para o planejamento e condução de entrevistas. Deve-se, portanto:

- Reunir pensamentos enquanto o participante preenche o questionário pós-teste.
- Rever o questionário pós-teste.
- Gravar em áudio a entrevista.
- Começar deixando o participante dizer tudo aquilo que lhe vem à cabeça.
- Começar as perguntas a partir dos tópicos de nível geral e mais elevado.
- Exemplo: questões de navegação, *layout* geral e design.
- Passar aos tópicos específicos.
- Rever aqueles locais no questionário pós-teste do participante que foram assinalados previamente como áreas a serem exploradas.
- Focalizar na compreensão de problemas e dificuldades, e não na solução de problemas.
- Terminar todas as perguntas antes de abrir o campo para a discussão pelos observadores.
- Caso seja apropriado, deixar aberta uma porta para contatos subsequentes.

É possível, também, utilizar a técnica denominada “Refazer o teste”. Essa técnica possui dois métodos: o método manual, que consiste em encontrar nos dados colhidos pelos participantes os pontos críticos e questioná-los sobre esses; o método em vídeo, que consiste em assistir o vídeo da realização da tarefa junto com o participante e questioná-lo sobre os pontos críticos.

Outra técnica que pode ser utilizada é a técnica “o que você lembra”. Ela consiste em mostrar ao participante, durante a entrevista, a tela com o elemento em questão removido ou recoberto, e então é perguntado ao participante se ele lembra o que tinha naquele local.

Nos casos em que o participante parece relutante em criticar o produto, embora tenha tido dificuldade em usá-lo, sugere-se

que o pesquisador mencione que está surpreso com a discrepância entre o desempenho e os comentários do participante ou então mencione que outros participantes fizeram críticas bastante opostas aos comentários do participante e que você precisa ter certeza absoluta sobre o que ele acha.

9.9 Análise dos dados dos testes de usabilidade

De acordo com Travis (2003), os testes de usabilidade podem ser analisados em termos de eficácia e satisfação. Pode-se também incluir alguma análise quantitativa como, por exemplo, uma escala de notas para satisfação ou uma escala para os comentários efusivos dos participantes.

Segundo Fernandez (2005), testes de usabilidade realizados em interfaces de *websites* tendem a oscilar entre dois pólos: ora com uma visão muito de observar os aspectos de design, ora enfatizando somente o lado de performance técnica ou mecânica da transação. É preciso equilibrar essas duas visões e ser extremamente detalhista no levantamento de problemas em ambas.

De acordo com Badre (2002), a usabilidade tem como mediadas os seguintes fatores:

- Tempo de aprendizagem.
- Tempo de execução.
- Número e tipos de erros.
- Facilidade de relembrar.
- Satisfação, preferências, opiniões e atitudes (medidas subjetivas).

Segundo Dumas e Redish (1999), a análise dos dados em qualquer teste de usabilidade sempre começa com a descrição das características dos dados. Há medidas de:

- Freqüência dos resultados, tais como o número de erros que ocorrem na tarefa.

- Resultados específicos, como a média ou mediana do tempo de uma tarefa.
- Grau de variabilidade dos resultados, tais como a faixa de tempo para a realização da tarefa.

Cabe ressaltar que muitas incorreções encontradas nos testes de usabilidade são sintomas de outros problemas maiores e não se constituem exatamente no centro da questão.

Segundo Travis (2003), é possível fazer análise rudimentar do tipo de problema que se identificou, como também é possível estimar o grau de severidade que acarretará na usabilidade - baixo, médio, sério ou crítico e montar uma tabela com os problemas de cada categoria.

Segundo Dumas e Redish (1999), um teste de usabilidade gera uma quantidade substancial de dados.

- Uma lista de problemas que tem crescido durante a produção do teste.
- Dados quantitativos sobre tempos, erros e outras medidas de performances;
- Dados quantitativos nos resultados subjetivos e outras perguntas a partir dos questionários pós-tarefa e pós-teste.
- Comentários dos participantes sobre seus *logs* ou notas e sobre os questionários.
- As notas escritas pela equipe do teste e seus comentários que tenham sido registrados num *log*.
- Dados sobre o histórico dos participantes, sobre os perfis dos usuários, questionário de recrutamento ou questionário pré-teste.
- *Videotapes* eventualmente mostrando diversas cenas do teste.

Colher os dados se faz necessário, mas não é o suficiente para os testes de usabilidade. Depois dos testes em si, é preciso

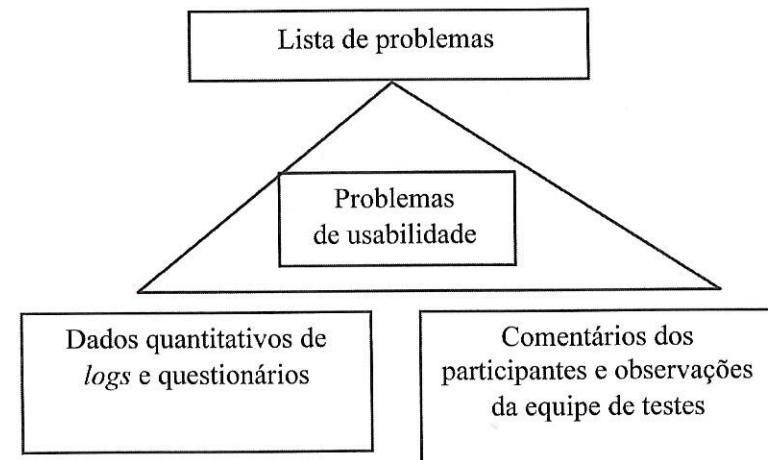
analisar os dados, considerando os quantitativos e qualitativos dos participantes, junto com as próprias observações e comentários dos usuários, de modo a usar todas essas informações para diagnosticar e documentar os problemas de usabilidade do produto e para recomendar soluções para esse problema. A ênfase do teste de usabilidade é encontrar problemas reais do produto e do processo que foram usados para o seu desenvolvimento.

Segundo Preece, Rogers e Sharp (2005), uma vez que a maioria dos testes com usuários envolve um número pequeno de participantes, somente estatísticas descritivas simples podem ser utilizadas para apresentar as descobertas: máximo, mínimo, média para o grupo e, algumas vezes o desvio-padrão.

Nos testes de usabilidade, há geralmente uma grande massa de dados a partir de dados sobre um pequeno número de participantes. Uma técnica muito útil para manusear uma grande quantidade de dados oriundos de diversas fontes é a triangulação. É preciso olhar todos juntos e, então, enxergar como cada parte dos dados se relaciona com os demais.

Se o pesquisador tiver desenvolvido uma lista de problemas durante o teste, provavelmente encontrará aquilo que possa vir a ser uma chave útil para análise. Isso se dá porque nós a colocamos no vértice do triângulo. O pesquisador deve, entretanto, estar aberto e alerta para a informação que as suas outras fontes revelam. É possível utilizar outras medidas quantitativas, notas, comentários dos participantes e sua lista de problemas para triangular os problemas encontrados.

A seguir, é apresentado um diagrama representativo de uma triangulação, proposto por Dumas e Redish (1999), usando múltiplas fontes de dados para encontrar problemas de usabilidade em um produto.



Esquema 14 – “Triangulação” usando múltiplas fontes de dados para encontrar problemas de usabilidade

9.10 Tabulação e manipulação dos dados

O primeiro passo para manipular os dados obtidos com os testes de usabilidade é fazer uma planilha resumida da medida de desempenho e de aspectos subjetivos. Para a medida do desempenho, será preciso começar tabulando-os para cada tarefa e participante. Por exemplo, será feita uma tabulação típica para cada tarefa, tempo decorrido, número de erros de vários tipos e outras medidas importantes como o número necessário de observações dos participantes.

Além do tempo para cada tarefa, devem ser computados o tempo total e o tempo médio. Os tempos são tabulados a partir de um software gerador de *log*. Mas é possível também usar as informações sobre tempo coletadas a partir de um cronômetro. Inserem-se, então, os tempos dentro da planilha de um computador pessoal e adicionam-se fórmulas para computar os totais e as médias.

Softwares de planilhas permitem computar rapidamente estatísticas simples e são aceitos como entrada pela maioria dos pacotes estatísticos complexos, bem como pelos processadores de texto (Dumas e Redish, 1999).

Criam-se também planilhas separadas para os itens do questionário pré-teste e para itens dos questionários pós-tarefa e pós-teste. O pré-teste fornece um perfil demográfico dos participantes. O pós-tarefa e o pós-teste fornecem as medidas subjetivas que serão tabuladas.

Além disso, deve-se fazer uma lista em um arquivo de texto com os comentários importantes feitos pelos participantes. Esse comentários são retirados dos campos de comentários nos questionários e de impressões dos dados dos *logs*. Finalmente, é gerada uma lista de comentários que os membros da equipe do teste fizeram durante o mesmo. Alguns são obtidos da planilha e outros do *log* de dados.

9.11 A estatística na análise dos testes de usabilidade

Segundo Nielsen (2004), há dois tipos principais de pesquisa sobre usuário: quantitativo (estatístico) e qualitativo (visões). O quantitativo tem vantagens interessantes, mas o qualitativo traz melhores resultados com menor custo, além dos benefícios de reduzirem uma situação complexa num número único que é fácil de obter e discutir.

Todo teste de usabilidade precisa de informação quantitativa. Para substanciar os problemas relatados, consideramos que serão tomadas ao menos algumas medidas básicas, tais como: o número de participantes que têm o problema, o número de escolhas erradas ou o tempo necessário para completar a tarefa. As medidas reais dependerão dos seus objetivos específicos e da

fase do projeto ou desenvolvimento na qual o teste está sendo realizado. As medidas podem vir de observações, de registros com um programa de captura de tela, ou da revisão de um *videotape* depois do teste.

Dumas e Redish (1999) ainda colocam que só é preciso usar estatísticas descritivas simples, tais como comentários dos participantes do teste, para documentar o caso com relação à presença de problemas de usabilidade, na maioria dos testes.

Nielsen (2004) demonstra que o fetichismo dos números leva estudos de usabilidade a se perderem ao se prenderem em análises estatísticas que são geralmente falsas, enviesadas, perturbadoras ou extremamente limitadas. É melhor enfatizar os *insights* e a pesquisa qualitativa.

9.12 Relatórios do teste de usabilidade

As observações do teste de usabilidade, aliadas às respostas das entrevistas com os participantes são transformadas em um relatório detalhado, contendo:

a) Descrição e metodologia do roteiro do teste

Uma breve descrição explicativa das metas propostas e de como foram alcançadas.

b) Análise do comportamento dos participantes

Como os participantes se comportam durante o teste. Parece irrelevante, mas isso é essencial para avaliar o sucesso da interface.

c) Resumos das entrevistas pós-teste

As principais opiniões, perguntas, dúvidas, reclamações e sugestões dadas pelos participantes após o teste.

d) Problemas de usabilidade

Lista dos problemas de usabilidade encontrados. Cada problema é tratado detalhadamente, com descrição dos critérios heurísticos que ele viola, seu grau de severidade, o custo para o cliente de não consertar o problema *versus* o custo de resolvê-lo.

e) Principais paradas críticas

Mais do que um simples problema, uma parada crítica é um beco sem saída do *site*. Seja físico (o usuário não tem mais fisicamente para onde ir) ou psicológico (o usuário fica tão confuso e frustrado que simplesmente pára, desiste). As paradas críticas, se existirem, têm que ser tratadas como prioridade absoluta.

f) Plano de correção

Cada solução deve sempre ser pensada em termos de custo-benefício para o cliente. Alguns problemas de usabilidade pouco severos, mas de solução custosa, podem ser ignorados, ao mesmo tempo em que problemas graves, e de soluções caras, terão que ser solucionados.

1.13 Exemplo de teste de usabilidade

Em um estudo sobre a usabilidade em formulários eletrônicos em *sites* de *e-commerce* (Santa-Rosa e Moraes, 2006b), foram avaliados quatro *sites* brasileiros de comércio eletrônico: *Submarino.com*, *Americanas.com*, *Ponto Frio* e *Casa & Vídeo*.



Figura 23 – Site: Americanas.com



Figura 24 – Site: Casaevideo.com.br



Figura 25 – Site: Submarino.com

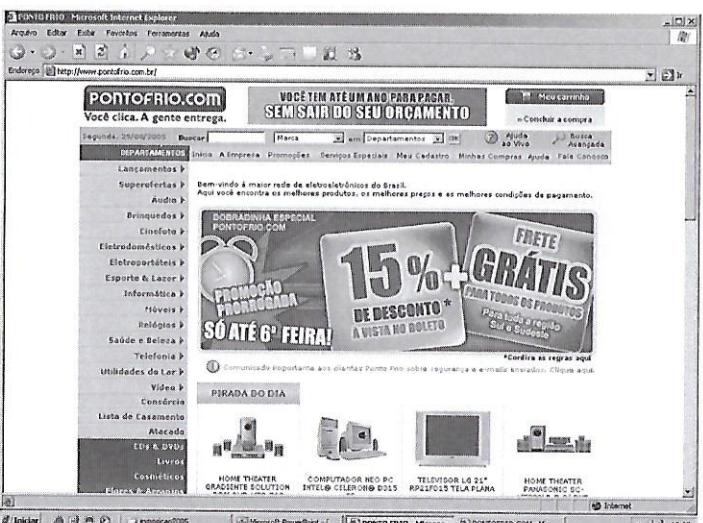


Figura 26 – Site: Pontofrio.com

Cabe ressaltar que o objetivo deste estudo não era a comparação e a determinação dos melhores ou piores *sites*, e sim levantar os problemas de usabilidade e, posteriormente, gerar diretrizes e recomendações para a elaboração de formulários para a venda *online*. Foram, então, selecionados e distribuídos 24 participantes, de acordo com as informações assinaladas em um quadro de horários preferenciais, e subdivididos aleatoriamente em grupos de seis participantes para cada *site* avaliado.

Os testes de usabilidade foram realizados em uma sala silenciosa. As janelas eram revestidas por película foto cromática, com o intuito de minimizar possíveis reflexos na tela do computador e na câmera de vídeo.

Para garantir o sigilo das informações dos participantes e diminuir o viés que o preenchimento de diferentes informações poderia causar, foi criada uma ficha com dados pessoais de um comprador fictício.

Cada participante, então, recebeu como tarefa a compra de um *disc-man* para presentear sua irmã. Destacam-se, como subtarefas, que os participantes deveriam especificar o endereço de entrega e se possível, solicitar embalagem para presente.

Os testes foram registrados em vídeo, e as interações capturadas pelo software de captura *Camtasia Studio*. Foram gravados também, os comentários tecidos e as perguntas feitas pelos participantes durante a realização dos testes.

Após a realização dos testes de usabilidade, os dados foram tabulados e então representados sob a forma do quadro: Tipos de problemas de usabilidade encontrados.

Problemas encontrados	Submarino	Americanas	Ponto Frio	Casa & Vídeo
Problemas na entrada de dados				
Problemas de navegação				
Problemas na entrada como novo usuário				
Problemas com endereço de entrega				
Problemas com campos obrigatórios				
Problemas com validação				
Problemas com rótulos dos campos				
Problemas na compreensão das instruções				
Problemas com instruções não pertinentes				
Problemas no reconhecimento de botões				
Problemas com disposição dos elementos				

Muitos problemas
Vários problemas
Poucos problemas
Nenhum problema

Quadro 5 – Tipos de problemas de usabilidade

Percebe-se que os problemas que apareceram com maior freqüência durante as sessões do teste de usabilidade foram, também, de modo geral, os mais críticos. A especificação do endereço de entrega e a validação geralmente feita de forma confusa gera insatisfação no usuário, atraso na realização da tarefa e, em alguns casos, a desistência.

No caso do *site* da Casa & Vídeo, por exemplo, alguns usuários demoraram tempo significativo para entrar, em função do endereço do *site* não utilizar o “& comercial”. Neste caso, uma barreira existe até mesmo antes de o usuário entrar no *site*.

Em alguns *sites* as dificuldades estavam no cadastro de novos usuários. Questões relativas ao posicionamento dos campos, botões e rótulos faziam com que os usuários se confundissem e prenchessem o seu e-mail no campo de usuário antigo.

Ficou evidente após a realização dos testes de usabilidade a necessidade de utilizar rótulos com terminologia, cores, posicionamento e tipografias adequadas.

Percebeu-se, também, que, de modo geral, os usuários com menor experiência dizem que os *sites* são muito bons e eficazes, mesmo que não tenham conseguido completar a tarefa. Eles atribuem o insucesso a si mesmos e não ao projeto do *site*.

Diferentemente das lojas físicas, onde em geral, os compradores levam mais tempo na busca e escolha dos produtos (caminhando pela loja à procura do produto desejado) – nos *sites* de comércio eletrônico verificou-se que o preenchimento dos formulários para cadastro e compra demorou mais na fase de *checkout* do que na seleção do produto.

Cabe enfatizar que os participantes, em sua maioria, procuram o produto em primeiro lugar por meio da navegação do *site*. Somente quando encontram dificuldades em achá-lo recorrem, então, ao campo de busca no *site*. Isso indica que o projeto de *sites* deve considerar os aspectos de arquitetura da informação de modo a criar classificações, menus hierárquicos e rótulos mais de acordo como entendimento e compreensão dos usuários.

Notou-se também carência de mecanismos para comparações internas, indexadas, tanto de preços quanto de especificações técnicas.

Pode-se observar, neste estudo, a necessidade de considerar o uso do mouse, do teclado, da leitura na tela, das barras de rolagem, dos rótulos e botões de navegação ao elaborar interfaces para formulários em comércio eletrônico.

Apesar de a usabilidade nos últimos anos ter sido bastante difundida e dos projetistas estarem mais conscientizados da importância do desenvolvimento de projetos centrados no usuário, continuamos encontrando problemas de usabilidade, como se pode observar nas imagens a seguir.



Figura 27 – Site: Americanas.com em 2007: Frete grátis é um botão



Figura 28 – Site: Americanas.com em 2007: Nesta outra imagem frete grátis está desabilitado



Figura 29 – Site: Casaevideo.com.br: Elemento frete grátis não se comporta como botão

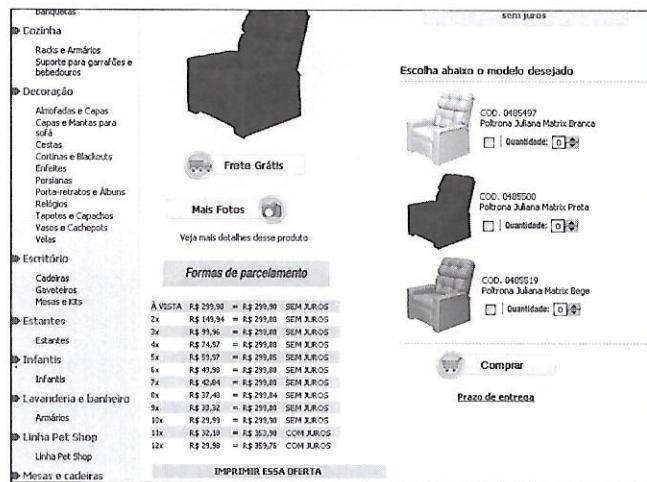


Figura 30 – Site Casaevideo.com.br: elemento “mais fotos” tem a mesma aparência de “frete grátis”, mas ambos se comportam como botões

Nota-se que, em ambos os *sites*, os elementos gráficos relativos ao “frete grátis” têm aparência de botões. No caso do *site* da Americanas.com, o “frete grátis” funciona como botão e na próxima página não. Claro que podemos considerá-lo como um botão desabilitado, mas, neste caso, seria melhor que o designer tivesse projetado alguma alteração visual, de modo que ficasse claro para os usuários que, nesta tela, o botão não teria função.

No caso do botão de “frete grátis” da Casa &Vídeo, o problema é que o frete grátis, segundo testes com usuários, tem aparência de botão e não se comporta como tal. Como se não bastasse, outros elementos têm aspecto visual semelhante ao frete grátis, mas se comportam como botões. A falta de consistência entre os elementos da interface pode ocasionar falhas e problemas na interação e gerar insatisfação do usuário.



Figura 31 – Menu do site casaevideo.com.br em 2007



Figura 32 – Menu do site: Americanas.com em 2007

Contudo, observando os menus dos sites Americanas.com e da Casaevideo.com.br em 2007, percebe-se que foi dada uma atenção maior às questões relativas à arquitetura da informação.

No caso da Casa & Vídeo foram criados pictogramas e associadas cores específicas para cada categoria de produtos. A Americanas.com, por sua vez, reduziu a densidade informacional do menu e incluiu a opção para ver *links* para os seus 34 departamentos de uma vez só.