

UNIVERSIDADE ESTADUAL DE MARINGÁ

SELMA ELAINE MARQUES HIDAE SCOMPARI

**DEFINIÇÃO DE REQUISITOS PARA MONTAGEM DE
UM LABORATÓRIO EXPERIMENTAL DE
AVALIAÇÃO DE INTERAÇÃO HUMANO
COMPUTADOR**

Maringá

2007

SELMA ELAINE MARQUES HIDAE SCOMPARIN

DEFINIÇÃO DE REQUISITOS PARA MONTAGEM DE UM LABORATÓRIO EXPERIMENTAL DE AVALIAÇÃO DE INTERAÇÃO HUMANO COMPUTADOR

Trabalho de Conclusão de Curso apresentado como parte dos requisitos para obtenção parcial do título de Especialista, do Curso de Especialização em Desenvolvimento de Sistemas para a *Web*, da Universidade Estadual de Maringá.

Orientador:
Professor Dr^o. Dante Alves Medeiros Filho

Maringá

2007

SELMA ELAINE MARQUES HIDAE SCOMPARIN

**DEFINIÇÃO DE REQUISITOS PARA MONTAGEM DE
UM LABORATÓRIO EXPERIMENTAL DE
AVALIAÇÃO DE INTERAÇÃO HUMANO
COMPUTADOR**

Trabalho de Conclusão de Curso apresentado como parte dos requisitos para obtenção parcial do título de Especialista, do Curso de Especialização em Desenvolvimento de Sistemas para a Web, da Universidade Estadual de Maringá.

Aprovado em ____ de _____ de 2007

Prof

Prof

AGRADECIMENTOS

Em primeiro lugar a Deus por esta sempre presente em minha vida.

À minha família, pelo apoio incondicional.

Ao professor Dante Filho, pelo incentivo e pela paciência.

Aos meus colegas da pós-graduação, pelo companheirismo, os momentos de diversão e por ter me recebido sempre de braços abertos. Aos funcionários do departamento, pelo carinho e atenção.

Aos amigos, sempre apoiando e dando a força necessária quando eu precisei.

Finalmente, agradeço à BS2 internet, por ter me ajudado direta e indiretamente.

SUMÁRIO

1	INTRODUÇÃO	06
2	REVISÃO DA LITERATURA – AVALIAÇÃO DE INTERFACES - IHC	07
2.1	INTERPRETAÇÃO DAS INTERFACES	07
2.1.1	OBJETIVOS DA AVALIAÇÃO DE INTERAÇÃO.....	08
2.1.2	CONCEITOS E QUALIDADE DE USOS	10
2.2	CARACTERÍSTICA DOS MÉTODOS DE AVALIAÇÃO.....	15
2.2.1	AVALIAÇÃO DO CICLO DE DESIGN.....	15
2.2.2	TÉCNICAS DE COLETAS DE DADOS	16
2.3	TIPO DE DADOS COLETADOS.....	19
2.4	TIPO DE ANÁLISE	20
2.5	GUIA PARA O PLANEJAMENTO DE UMA AVALIAÇÃO.....	21
2.6	MÉTODOS DE AVALIAÇÃO ANALÍTICOS	22
2.6.1	AVALIAÇÃO HEURÍSTICA.....	25
2.6.2	PERCURSO COGNITIVO.....	30
3	MÉTODOS EXPERIMENTAIS	35
3.1	MÉTODOS DE AVALIAÇÃO EMPÍRICOS	35
3.2	COMPOSIÇÃO DO MATERIAL PARA O TESTE	39
3.3	TESTES EM LABORATÓRIOS	41
3.4	TESTES DE USABILIDADE	45
3.5	TESTES DE COMUNICABILIDADE	47
3.6	TESTES DE USABILIDADE X TESTES DE COMUNICABILIDADE	54
3.7	PROTOCOLOS VERBAIS.....	54
3.8	AVALIAÇÃO NO AMBIENTE DO USUÁRIO	56
4	REQUISITOS PARA MONTAGEM DO LABORATÓRIO	58
5	CONSIDERAÇÕES FINAIS	60
	REFERÊNCIAS BIBLIOGRÁFICAS	61

1 Introdução

Atualmente com o avanço da computação gráfica e da utilização de computadores nos mais diversos setores da vida humana, cresce o interesse pelo estudo do desenvolvimento e da avaliação de interfaces humano-computador. Nas universidades e nos centros de pesquisa existem muitas investigações nesta área procurando facilitar a utilização de computadores. Neste contexto uma das grandes linhas de investigação é a da avaliação de interfaces.

No presente trabalho procurou-se apresentar vários métodos de avaliação de interfaces humano computador com o objetivo de prover uma lista de equipamentos, materiais e recursos humanos necessários à implantação de um laboratório para avaliação experimental de interfaces.

Para o desenvolvimento dessa tarefa destinou-se neste trabalho um capítulo para revisão de literatura sobre métodos de avaliação com o intuito de subsidiar o levantamento do que é necessário para a montagem de um laboratório experimental destinado a avaliação de interfaces.

2 Revisão de Literatura - Avaliação de Interfaces - IHC

Toda avaliação é traçada por métodos, modelos e diretrizes. Os estudos sobre avaliação de IHC são relacionados como estruturar as interfaces com a ergonomia, usabilidade e comunicabilidade, durante o processo de desenvolvimento quanto em sua finalização.

2.1 Interpretação das Interfaces

A forma que o usuário interpreta o que está visualizando, possibilita a evolução em sua utilização na web. Interação é o processo de comunicação entre pessoas e sistemas interativos (Preece et al., 1994). A IHC estuda cada ação que o usuário venha tomar na utilização do sistema, e como foram suas reações nas interpretações através das interfaces (Figura 1).

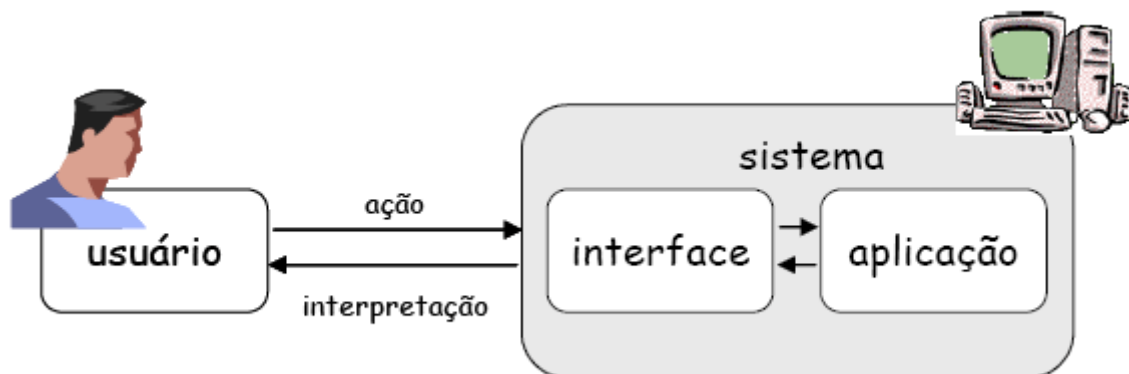


Figura 1: O processo de interação humano-computador.
(Fonte: Preece et al., 1994)

A **interface** é denominado a toda a porção de um sistema que o usuário mantém contato quando utilizado, tanto na forma passiva ou ativa. A interface engloba tanto

software quanto hardware (dispositivos de entrada e saída; teclados, mouse, tablets, monitores, impressoras e etc.) Todo processo de interação responde a uma comunicação, assim sendo a interface pode ser vista como um sistema de comunicação utilizado neste processo. Uma definição de interface por Moran:

“a interface de usuário deve ser entendida como sendo a parte de um sistema computacional com a qual uma pessoa entra em contato — física, perceptiva ou conceitualmente” (Moran, 1981).

Podemos dizer que há 3 tipos de dimensões a física, perceptiva e conceitual. A física consiste no conjunto de interface que o usuário pode manipular, perceptiva o todo que o usuário possa perceber. A conceitual é a forma que o usuário interpreta e raciocina com a interação no sistema, com base em suas características físicas e cognitivas no seu cotidiano.

As interfaces atuais são mais comuns através de elementos visuais e sonoros, com a entrada de dados via teclado e mouse.

2.1.1 Objetivos da Avaliação de Interação

Mesmo que o sistema esteja finalizado, é importante fazer avaliação se esta adequado com os usuários em suas utilizações e ambiente em que será utilizado. Os testes de funcionalidade são imprescindível para verificar se houve conflito ou redundância de informação, a avaliação de interface é necessária para se analisar a qualidade de uso de um software. Quanto mais cedo forem encontrados os

problemas de interação ou de interface, menor o custo de se consertá-los (Karat, 1993).

O sistema finalizado sempre estará em avaliação, nem que seja apenas um usuário.

Alguns dos principais objetivos de se realizar avaliação de sistemas interativos são (Hartson, 1998; Preece et al., 2002):

- identificar as necessidades de usuários ou verificar o entendimento dos projetistas sobre estas necessidades
- identificar problemas de interação ou de interface
- investigar como uma interface afeta a forma de trabalhar dos usuários
- comparar alternativas de projeto de interface
- alcançar objetivos quantificáveis em métricas de usabilidade
- verificar conformidade com um padrão ou conjunto de heurísticas

Mas na maiorias das vezes os gerentes de projetos vetam a realização das avaliações em seus sistemas por envolver custo. A avaliação IHC deveria ser como um IMETRO um selo de qualidade e confiabilidade, por avaliar um sistema em sua execução podendo agregar economia de tempo, por visualizar uma forma de implantação de interface mas eficaz .

2.1.2 Conceitos e qualidade de usos

Tudo que venha a facilitar ao usuário interagir com eficiência e satisfação. A utilização de modo confiável e seguro, passa ao usuário realizar suas tarefas no sistema com mais tranqüilidade e espontaneidade, dependendo em grande parte da qualidade de uso do sistema. O conceito de qualidade de uso mais amplamente utilizado é o de usabilidade, relacionado à facilidade e eficiência de aprendizado e de uso, bem como satisfação do usuário (Nielsen, 1993). Mais recentemente, foi elaborado o conceito de comunicabilidade, que busca avaliar o processo implícito de comunicação designer– usuário, que se dá através da interface (Prates et al. 2000b, Prates et al., 2000a). Já o conceito de aplicabilidade está relacionado à flexibilidade de um sistema, em particular com relação à sua utilidade em uma variedade de situações (Fischer, 1998).

Usabilidade

Consiste na facilidade de uso do sistema e na sua produtividade. Alguns fatores típicos envolvidos no conceito de usabilidade são (Nielsen, 1993; Preece et al., 2002):

- facilidade de aprendizado
- facilidade de uso
- eficiência de uso e produtividade
- satisfação do usuário
- flexibilidade
- utilidade

- segurança no uso

Facilidade de aprendizado é a eficiência de que forma foram colocadas as informações através das interfaces que o usuário possa aprender com competência e desempenho. Quando um sistema é analisado de forma que o uso seja simples, considerando um nível intermediário ou avançado, sendo que cada tipos e graus de aprendizado distintos.

Facilidade de uso em um sistema não só apenas o cognitivo apresentado no sistema, e sim os erros cometidos no processo de interação. É importante observar que um sistema nem sempre é fácil de aprender não é necessariamente fácil de utilizar ou vice-versa.

Eficiência de uso, é verificar se o sistema faz o que ele propõem a fazer e **produtividade** pode ser verificada se o sistema faz com que o usuário execute o que propõem de forma rápida e eficaz. A avaliação é feito pelo tempo decorrido desde do início até a finalização que o usuário se propôs a realizar.

Satisfação do usuário, consiste na avaliação subjetiva que leva o usuário a cada reação emocional que possa surgir durante a interação, sejam elas positivas, como prazer e diversos, ou negativas, como tédio ou frustração.

Flexibilidade, é a maneira que o usuário reage na utilização do sistema, e o quanto um sistema pode disponibilizar.

Utilidade, no sistema refere-se um conjunto de funcionalidades necessárias para que o usuário pode realiza. Esta dimensão está intimamente relacionada ao conceito de aplicabilidade proposto por Fischer (1998).

Segurança no uso, se refere ao grau de proteção de um sistema contra condições desfavoráveis ou até mesmo perigosas para os usuários. Trata-se principalmente de como evitar e permitir que o usuário se recupere de condições de erro com conseqüências sérias para seu trabalho ou para sua saúde.

Comunicabilidade

O conceito de **comunicabilidade** (de Souza et al. 1999, Prates et al., 2000b) se refere à capacidade de os usuários entenderem o design tal como concebido pelos projetistas. A hipótese subjacente ao conceito de comunicabilidade é que, se um usuário entende as decisões que o projetista tomou ao construir a interface, aumentam suas chances de fazer um bom uso daquele sistema. Em sistemas com alta comunicabilidade, os usuários são capazes de responder:

- para que o sistema serve
- qual é a vantagem de utilizá-lo
- como funciona
- quais são os princípios gerais de interação com o sistema

Para que o sistema seja desenvolvido com a eficaz e seus objetivos que o sistema requer, é necessário que um briefing seja bem formulado, assim o designer possa elucidar com clareza a necessidade do sistema e transferir para interface. Toda

interface para que seja boa em sua comunicabilidade é buscar o cognitivo e o cotidiano que o interprete que é o usuário possa interagir com maior facilidade por que estará com símbolo (ícones) representando o mundo que ele se familiariza.

Aplicabilidade

A **aplicabilidade** de um sistema também determina sua qualidade de uso. Este conceito está relacionado com a utilidade deste sistema em uma variedade de situações e problemas (Fischer, 1998).

Fischer citar em suas literatura que as pessoas, por se acharem hábeis e especialistas no seu domínio querem se posicionar como designers, para solucionar problema de construção ou transformação dos seus próprios artefatos e espaço de trabalho. Para ele o IHC tem que ser desafiado, para melhoria das formas como as pessoas podem usar o PC para trabalharem, pensarem, se comunicarem, aprenderem, criticarem, explicarem, argumentarem, discutirem, observarem, decidirem, calcularem, simularem e projetarem.

Interfaces difíceis e rígidas, força o usuário a utilizar o sistema somente da mesma maneira, sem o mesmo a vir cometer algum erros. Esse tipo de interface leva ao usuário a pensar que ele não é capaz de tomar decisões apropriadas. A probabilidade de erro do usuário nesse sistema com essa interface são baixos, mas isso não é ponto favorável ao sistema por que leva o usuário a se afastar por não fazer o mesmo a pensar.

Diversos pesquisadores têm chamado atenção para a necessidade de se desenvolver sistemas que **ampliem as capacidades dos usuários**, em vez de tentarem substituí-las, possibilitando que eles ajam de forma mais inteligente e eficiente (Adler & Winograd, 1992).

Interfaces com baixa qualidade de uso trazem diversos problemas, dentre os quais:

- requerem treinamento excessivo
- desmotivam a exploração
- confundem os usuários
- induzem os usuários ao erro
- geram insatisfação
- diminuem a produtividade
- não trazem o retorno de investimento previsto
-

Durante o processo de avaliação, são usados diversos métodos que podem detectar estes problemas ao longo do desenvolvimento. Os métodos de avaliação mais usados são os de **usabilidade** e a **comunicabilidade** de um sistema. No conceito aplicabilidade não tem um método específico para sua avaliação, assim se opta por um dos métodos qualitativos de avaliação, que possibilita captar informações valiosas para esta avaliação.

2.2 Característica dos métodos de avaliação

Cada método de avaliação se diferencia por sua forma de avaliar específico comportamento das interfaces com os usuários. As principais diferenças entre os métodos são a etapa do ciclo de design do sistema em que devem ou podem ser aplicados (durante o ciclo de desenvolvimento ou após ter o produto pronto), a técnica utilizada para coletar os dados (desde entrevistas até experimentos em laboratórios), os tipos de dados coletados (quantitativos ou qualitativos), e ainda o tipo de análise feito (o avaliador pode prever potenciais problemas ou interpretar os dados obtidos) (Preece et al., 1994).

2.2.1 Avaliação do ciclo de design

Durante as etapas do ciclo de desenvolvimento da interface pode ser feita a avaliação do sistema. As avaliações formativas (Preece et al., 1994; Hartson, 1998) são aquelas que são feitas durante o processo de design, antes de o sistema estar terminado, e muitas vezes antes de uma linha de código sequer ser escrita. A avaliação pode ser feita utilizando-se desde cenários, *storyboards*, ou a modelagem conceitual da interação até protótipos do sistema. A vantagem de se fazer avaliação formativa é que problemas de interação são identificados e consertados antes de a aplicação ser terminada e liberada para uso. Quanto mais cedo no ciclo de design um problema é descoberto e reparado, menor o custo das alterações necessárias no software (Karat, 1993). Tendo um resultado no sistema com melhor qualidade ao usuário.

Todas as avaliações feitas em produtos já finalizados são denominados como avaliação somativas. Normalmente, enquanto as avaliações formativas têm por objetivo melhorar a qualidade do sistema, tornando-o mais usável para o usuário, as avaliações somativas buscam verificar a existência de determinados aspectos no sistema desenvolvido, como por exemplo, a sua conformidade com um padrão estabelecido.

2.2.2 Técnicas de Coletas de Dados

Há varias técnicas de coletar dados sobre uma interface de um sistema e se fazer uma análise de qualidade de uso. A decisão sobre que técnica utilizar depende principalmente da disponibilidade dos recursos que se tem e objetivos da avaliação a ser feita.

Coleta da opinião de usuários

O objetivo da coleta da opinião de usuário é obter uma apreciação dos usuários em relação ao sistema interativo. Identificar o nível de satisfação dos usuários com o sistema, incluir certos aspectos como: se eles gostam do sistema, se a aparência estética do sistema é satisfatória, se o sistema faz aquilo que eles desejam, se tiveram algum problema ao usá-lo, e/ou se eles gostariam de (ou pretendem) usá-lo novamente. As principais técnicas utilizadas para se coletar a opinião de usuários são questionários e entrevistas. Os tipos de questionários e entrevistas a serem utilizados variam em diversos aspectos (Nicolaci-da-Costa, 2001; Preece et al., 2002).

As coletas da opinião podem ser feitas através de telefonemas, pessoalmente, email ou web, com um pequeno grupo de pessoas ou grupos maiores, e utilizando perguntas bem estruturadas ou espontâneas.

Observação de usuários

Os usuários nem sempre conseguem expor o conhecimento da sua experiência de uso com o sistema. A avaliação sendo observada permite ter uma visão como o usuário se comporta na utilização do sistema, sendo assim possibilita ver os problemas vivenciados pelos usuários, sendo muitas vezes um resultado positivo durante o uso do sistema.

A observação do sistema pode ser feita no seu trabalho, ou até mesmo em sua casa. Nestas situações se consegue observar o usuário utilizando o software para realizar as tarefas que ele considerar relevantes e para as quais acredita que o sistema seja apropriado, e ainda da maneira que considera adequada ou desejável. Neste tipo de avaliação são encontrados alguns desafios para os avaliadores, de conseguir uma observação sem interferir no ambiente comum do usuário, e como fazer a análise dos dados coletados, principalmente quando se obtém várias horas de vídeo gravadas, ou quando diferentes tipos de registro feitos durante o uso devem ser integrados.

A avaliação pode ser feita com o usuário em ambientes mais controlados, como laboratórios. No ambiente de laboratório o avaliador tem um controle maior sobre os fatores diretos e indiretos que o usuário teve como dificuldade ou rendimento, como

tempo da avaliação, concentração do mesmo na utilização ou nas tarefas pedidas para serem executadas. Desse forma é possível coletar dados mais precisos sobre o sistema na sua utilização de diferentes usuários de forma que poderá compara-los. Nestes casos, o avaliador normalmente determina as tarefas a serem executadas pelo usuário e pode coletar dados qualitativos ou quantitativos sobre o uso, como por exemplo, o tipo e frequência das dificuldades enfrentadas pelos usuários, caminhos preferenciais, o tempo gasto para executar a tarefa, ou o número de erros cometidos.

Registro de uso

Para que seja feito o registro de uso o usuário e o avaliador tem que estar presente no mesma hora e local para que seja efetuada a avaliação. Mas nem sempre é possível, ou se gostaria de ter informações sobre o uso em um período de tempo mais longo. Uma outra forma de coletar informações sobre como os usuários usam o sistema é através de registros feitos durante o uso. Isto pode ser feito através de logs, que armazenam em um arquivo as ações executadas em um sistema, através da gravação da interação do usuário com o sistema, ou da gravação em vídeo da experiência do usuário.

As diferentes formas de registro armazenam aspectos distintos do uso feito. Normalmente, registros de uso geram uma grande quantidade de informação e a análise destes dados pode ser bastante custosa para avaliadores.

Coleta da opinião de especialistas

Nem sempre há usuários disponíveis para participar da avaliação, ou a sua participação gera um custo elevado. Em situações assim, a avaliação pode ser feita sem a participação de um usuário. Neste caso, pode se considerar a coleta da opinião de especialistas em IHC e/ou no domínio da aplicação. Os especialistas examinam a interface e identificam possíveis dificuldades que os usuários podem vir a ter ao utilizar o software.

2.3 Tipo de Dados Coletados

Os dados coletados a partir de uma avaliação de interface podem ser quantitativos ou qualitativos. Dados quantitativos são aqueles que podem ser representados numericamente. Normalmente dados quantitativos são utilizados para se avaliar a eficiência e produtividade de um sistema, para se comparar alternativas de design ou ainda se determinar se o sistema atingiu algum objetivo de qualidade de uso prédefinido. A análise destes dados freqüentemente é feita utilizando-se cálculos estatísticos simples, como desvio padrão ou médias. Alguns exemplos de dados quantitativos são o número de erros ocorridos ou o tempo gasto para completar uma tarefa.

Dados qualitativos são resultados não numéricos, tais como uma lista de problemas que os usuários tiveram ao utilizar a aplicação, ou suas sugestões sobre como melhorar o projeto de interação. Normalmente estes dados permitem identificar quais são as características de interação ou interface relacionadas com os

problemas medidos e observados. Em alguns casos, dados qualitativos podem ser categorizados e então quantificados.

2.4 Tipo de Análise

Os avaliadores podem analisar os dados coletados de forma preditiva, interpretativa ou experimental. A **análise preditiva** quando os avaliadores, ao analisarem os dados coletados de especialistas, tentam diagnosticar que tipo de problemas os usuários enfrentarão. A análise pode ser feita através de uma inspeção da interface ou em função de técnicas de modelagem.

A **análise interpretativa** é ao analisarem os dados coletados a partir da interação do usuário com o sistema, os avaliadores procuram explicar os fenômenos que ocorreram durante esta interação. A análise é considerada como sendo interpretativa quando ela é feita sobre dados coletados em ambientes naturais sem interferência dos observadores nas atividades dos usuários.

Todos os dados coletados em ambientes controlados, como laboratórios, precisam ser analisados em função das variáveis sendo observadas. A análise da avaliação também depende da interpretação do avaliador, diferenciando-a da anterior, pelo fato de as variáveis sendo manipuladas serem conhecidas. Desta forma denominamos este tipo de análise de **experimental**.

2.5 Guia para o planejamento de uma avaliação

Para que as decisões sejam tomadas com respeito a uma avaliação de IHC, propõe-se utilizar o esquema DECIDE (Preece et al., 2002), para ajudar avaliadores inexperientes no planejamento e realização de uma avaliação. Os pontos relevantes a serem considerados são os seguintes:

- **Determinar os objetivos gerais que a avaliação deverá tratar.** Trata-se de responder as perguntas gerais: Quais são os objetivos gerais da avaliação? Quem quer realizá-la e por quê?
- **Explorar perguntas específicas a serem respondidas.** Trata-se de decompor as perguntas gerais em perguntas específicas ao sistema a ser avaliado, considerando-se os usuários-alvo e suas atividades. Estas perguntas são necessárias para permitir efetivamente que os objetivos da avaliação sejam atingidos.
- **Escolher (*Choose*) o paradigma e as técnicas de avaliação que responderão as perguntas.** Dentre os pontos a serem considerados, destacam-se o prazo, o custo, os equipamentos e o grau de conhecimento e experiência dos avaliadores exigidos por cada técnica.
- **Identificar questões práticas que devem ser tratadas.** Consideram-se aqui fatores como: perfil e número de usuários que participarão da avaliação; ambiente em que a avaliação será realizada; seleção das tarefas; planejamento e preparação do material de avaliação; alocação de pessoal, recursos e equipamentos para a realização da avaliação.
- **Decidir como lidar com questões éticas.** Quando uma avaliação envolve

peessoas como participantes de testes, os avaliadores devem se certificar que os direitos destas pessoas estão sendo respeitados. Na seção 3.1 apontamos formas de lidar com as questões éticas envolvidas neste tipo de avaliação.

- **Avaliar (*Evaluate*), interpretar e apresentar os dados.** Como apontado nesta seção, os dados coletados durante uma avaliação podem variar bastante. Sendo assim, é importante considerar aspectos como a confiabilidade dos dados (se a técnica produz os mesmos resultados nas mesmas circunstâncias), sua validade (se mede o que deveria medir); potenciais distorções; escopo (o quanto as descobertas podem ser generalizadas); e validade ecológica (o quanto o ambiente em que a avaliação é feita influencia ou distorce os resultados).

2.6 Métodos de Avaliação Analíticos

Métodos de avaliação analíticos são aqueles nos quais avaliadores inspecionam ou examinam aspectos de uma interface de usuário relacionados a usabilidade (Mack & Nielsen, 1994). Normalmente, os avaliadores são especialistas em usabilidade, mas também podem ser consultores de desenvolvimento de software especializados em um determinado estilo de interface, ou ainda usuários finais conhecedores do domínio e da tarefa, entre outros.

A **avaliação analítica** ou por inspeção é utilizada geralmente para buscar problemas de usabilidade em um projeto de interface existente, e analisar estes problemas com vistas a fazer recomendações para consertá-los e assim melhorar a usabilidade do

projeto. Mack e Nielsen (1994) identificam como principais objetivos deste tipo de avaliação:

- **identificação de problemas de usabilidade:** identificar, classificar e contar o número de problemas de usabilidade encontrados durante a inspeção;
- **seleção dos problemas que devem ser corrigidos:** após identificar os problemas, a equipe de projeto deve reprojeter a interface para corrigir o maior número possível de problemas. Os problemas a serem corrigidos são priorizados de acordo com a gravidade do problema e o custo associado à correção.

Existem três tipos de conhecimento envolvidos em uma avaliação analítica: conhecimento sobre o domínio, conhecimento e experiência no projeto e avaliação de interfaces de usuário, e experiência em se realizar um tipo específico de avaliação (Lynch & Palmiter, 2002).

O **conhecimento sobre o domínio** é necessário para determinar o que os usuários querem, do que eles precisam, quais são as tarefas mais freqüentes e quais as mais importantes.

O **conhecimento e a experiência de projeto de interfaces de usuário** são necessários para que o avaliador seja capaz de analisar os aspectos mais importantes de um projeto de interfaces (navegação, terminologia, estruturas de controle, etc.) e para que ele conheça os princípios e diretrizes disponíveis na literatura. Sendo que os princípios e diretrizes costumam ser bastante genéricos,

esta experiência é fundamental para que o avaliador saiba distinguir quando aplicar e quando ignorar um princípio ou diretriz.

A **experiência em realizar um tipo de avaliação específico** dá ao avaliador a capacidade de representar um cliente, tendo em vista sobre o que procurar e o que relatar como resultado da avaliação.

Nestes tipos de conhecimento, apresentado pode-se avaliar os perfis de possíveis avaliadores conforme a seguinte escala, em ordem de preferência:

- **ideal**: especialista “duplo”. Possui experiência tanto nos processos e princípios de usabilidade quanto nos processos e aspectos relevantes do domínio.
- **desejável**: especialista em IHC. Conhece o processo de avaliação, bem como os princípios e diretrizes relevantes. Pode aprender o suficiente do domínio para selecionar as tarefas que deverão ser realizadas.
- **menos desejável**: especialista no domínio. Tem conhecimento do domínio e estuda princípios de interface e o processo de avaliação para realizar a avaliação.
- **menos desejável ainda**: membro da equipe de desenvolvimento. Tem dificuldade em deixar de lado seu papel de desenvolvedor e assumir um ponto de vista semelhante ao de um usuário.

Dependendo do caso de domínio, ter um problema desconhecido pelos avaliadores, pode ser necessário incluir no planejamento da avaliação um especialista no domínio, para esclarecer eventuais dúvidas dos avaliadores. Esta ajuda pode tomar

a forma de criação de cenários de uso típicos, que listem vários passos que um usuário com determinado perfil deve seguir para realizar uma amostra realista de tarefas. Tais cenários devem ser construídos com base em uma análise das tarefas dos futuros usuários do sistema.

Existem diversos tipos de avaliação analítica: avaliação heurística, percurso cognitivo, percurso pluralista, conformidade com diretrizes e padrões, e inspeções de consistência, de características ou formais (Mack & Nielsen, 1994).

2.6.1 Avaliação Heurística

A avaliação heurística é um método analítico que visa identificar problemas de usabilidade conforme um conjunto de heurísticas ou diretrizes (*guidelines*) (Nielsen, 1994). Ele se baseia em melhores práticas definidas por profissionais experientes e especialistas em IHC, ao longo de diversos anos de trabalho nesta área.

O método avaliação heurística não precisa de usuários, e é realizado por avaliadores especialistas. Normalmente, recomenda-se que 3 a 5 especialistas realizem esse método. É bem rápido, e de menor custo que a maior parte dos métodos de avaliação amplamente difundidos.

Como em todo método de avaliação, a avaliação heurística envolve uma fase de preparação, na qual se definem:

- proposta de design (papel ou protótipo)
- hipóteses sobre os usuários (opcional)
- cenário de tarefas (opcional)

A avaliação segue o seguinte procedimento:

a. sessões curtas (1 a 2 horas) de avaliação individual, onde cada especialista...

- julga a conformidade da interface com um determinado conjunto de princípios (“heurísticas”) de usabilidade
- anota os problemas encontrados e sua localização
- julga a gravidade destes problemas
- gera um relatório individual com o resultado de sua avaliação e comentários adicionais

É necessário que o avaliador esteja sozinho para que não seja influenciado por outra opinião. Cada sessão de avaliação, o avaliador percorre a interface diversas vezes, inspecionando os diversos elementos de interface e comparando-os com a lista de heurísticas de usabilidade.

b. consolidação da avaliação dos especialistas

- novo julgamento sobre o conjunto global dos problemas encontrados
- relatório unificado de problemas de usabilidade

Na etapa 2, cada avaliador tem acesso aos relatórios individuais de todos os avaliadores, e pode expressar seu julgamento sobre os problemas apontados pelos outros avaliadores. Ao final desta etapa, deve-se gerar um relatório unificado e consolidado sobre todos os problemas encontrados.

c. seleção dos problemas que devem ser corrigidos

Nesta etapa deve ser realizada junto ao cliente ou ao gerente de projeto. É análise de custo/benefício das correções aos problemas encontrados na etapa anterior. Esta análise deve levar em conta não apenas a gravidade dos problemas, mas também os prazos e o orçamento do projeto, bem como a capacitação da equipe de desenvolvimento.

Nielsen propôs um conjunto básico de heurísticas (Nielsen, 1993). Cada elemento de interface (ou conjunto de elementos) deve ser analisado para verificar sua conformidade com cada uma das seguintes heurísticas:

- **visibilidade do estado do sistema:** mantenha os usuários informados sobre o que está acontecendo, através de *feedback* adequado e no tempo certo.
- **correspondência entre o sistema e o mundo real:** utilize conceitos, vocabulário e processos familiares aos usuários.
- **controle e liberdade do usuário:** forneça alternativas e “saídas de emergência”, possibilidades de *undo* e *redo*
- **consistência e padronização:** palavras, situações e ações semelhantes devem significar conceitos ou operações semelhantes; caso haja convenções para o ambiente ou plataforma escolhidos, estas devem ser obedecidas
- **prevenção de erro:** tente evitar que o erro aconteça, informando o usuário sobre as conseqüências de suas ações ou, se possível, impedindo ações que levariam a uma situação de erro
- **ajuda aos usuários para reconhecerem, diagnosticarem e se recuperarem de erros:** mensagens de erro em linguagem simples, sem

códigos, indicando precisamente o problema e sugerindo de forma construtiva um caminho remediador

- **reconhecimento** em vez de memorização: torne objetos, ações e opções visíveis e compreensíveis
- **flexibilidade e eficiência de uso**: ofereça aceleradores e caminhos alternativos para uma mesma tarefa; permita que os usuários customizem ações freqüentes
- **design estético e minimalista**: evite porções de informação irrelevantes. Cada unidade extra de informação em um diálogo compete com as unidades de informação relevantes e reduz sua visibilidade relativa.
- **ajuda e documentação** devem ser fáceis de buscar, focadas no domínio e na tarefa do usuário, e devem listar passos concretos a serem efetuados para atingir seus objetivos

Todo problema encontrado, ou seja, para cada heurística violada, deve-se definir ainda a localização do problema, ou seja, onde ele ocorre na interface, e sua gravidade.

Com relação à **localização**, o problema pode estar:

- em um único local na interface;
- em dois ou mais locais na interface, casualmente;
- na estrutura geral da interface, de forma sistemática; ou
- pode ser algo que “não está lá”, ou seja, precisa ser incluído na interface

A gravidade (severidade) do problema é somada, por cada especialista, como uma combinação dos fatores:

- frequência com que o problema ocorre: É um problema comum ou raro?
- impacto do problema: Será fácil ou difícil para os usuários superarem o problema?
- persistência do problema: É um problema que ocorre apenas uma vez e que os usuários conseguem superar facilmente, ou os usuários serão incomodados pelo problema repetidas vezes?

A gravidade do problema é definida por um valor da seguinte escala:

- 0 – Não concordo que isto seja um problema (este valor pode resultar da avaliação de um especialista sobre um problema apontado por outro especialista)
- 1 – Problema cosmético: não precisa ser consertado a menos que haja tempo extra no projeto
- 2 – Problema pequeno: o conserto deste problema é desejável, mas deve receber baixa prioridade
- 3 – Problema grande: importante de ser consertado; deve receber alta prioridade
- 4 – Catastrófico: é imperativo consertar este problema antes do lançamento do produto

Na avaliação heurística, os especialistas fazem um relatório consolidado. Neste relatório pode conter, por exemplo, os seguintes itens:

- problemas esperados (e possíveis consertos)
- o quão bem o sistema apóia as tarefas dos usuários
- caminhos de interação primários (importantes e/ou freqüentes)
- caminhos de interação alternativos ou pouco utilizados
- consistência
- elementos de estilo
- recomendações de projeto

A avaliação heurística pode ser realizada no início da etapa de desenvolvimento. Quando o projeto ainda esta no papel, sem ter sido implementado no sistema as interfaces.

2.6.2 Percurso Cognitivo

O percurso cognitivo é um método analítico que avalia uma proposta de projeto de IHC no contexto de tarefas específicas do usuário (Wharton et al., 1994). Ele reforça que avaliar principalmente a facilidade de aprendizado do sistema, em particular pela exploração dos usuários. A motivação para este tipo de avaliação sobrevir do fato de que muitas pessoas priorizam em aprender sobre a funcionalidade de um sistema computacional enquanto trabalham em suas tarefas típicas, adquirindo conhecimento sobre novas características ou funções apenas quando seu trabalho as requerer.

O método cognitivo, requer que o aprendizado deve ser determinado pelo benefício imediato aos seus usuários. Os principais itens a ser investigado:

- a correspondência entre a conceitualização de uma tarefa por parte dos usuários e dos designers
- escolha adequada (ou inadequada) de termos, ou seja, o vocabulário utilizado
- *feedback* adequado (ou inadequado) para as conseqüências de uma ação

A hipótese subjacente a este método é que, num bom projeto de interface, as intenções dos usuários causam a seleção da ação adequada. Caso isto não aconteça, são levantadas hipóteses sobre as possíveis causas dos problemas, para que sejam estudadas e propostas soluções alternativas.

Não necessariamente que o processo cognitivo envolva o usuário. Pode realizar este processo individualmente, pelo próprio projetista que mostra uma nova solução, ou em grupo. O grupo pode compor de um projetista, a equipe do projeto, o marketing e de documentação, além de especialistas em avaliação de interfaces.

Antes de realizar a avaliação, deve-se passar por uma fase de preparação, na qual se definem:

- hipóteses sobre os usuários e sobre o conhecimento que eles têm sobre a tarefa e sobre a interface proposta

- cenários de tarefas, construídos a partir de uma seleção de tarefas importantes e de tarefas freqüentes
- seqüência “correta” de ações para completar cada tarefa, tal como definida pelo projetista
- proposta de design em papel ou protótipo, ilustrando cada passo e indicando o estado da interface antes/depois de cada um

Passos de execução da avaliação são os seguintes:

- o projetista apresenta uma proposta de design
- os avaliadores constroem histórias plausíveis sobre a interação de um usuário típico com a interface, com base nos cenários de tarefas selecionados
- os avaliadores simulam a execução da tarefa, efetuando uma série de perguntas sobre cada passo
- os avaliadores anotam pontos-chave, como:
 - ✓ o que o usuário precisa saber antes de realizar a tarefa
 - ✓ o que o usuário deve aprender ao realizar a tarefa

Cada evolução do projeto, os avaliadores fazem uma serie de perguntas, visando descobrir problemas em potencial, que poderiam ocorrer durante a interação com os usuários finais na utilização do sistema. Uma lista de perguntas básicas, bem como a esclarecer pontos específicos associadas a cada um:

- O usuário tentará atingir a meta correta?

- • Dada a decomposição de uma tarefa em subtarefas, o usuário saberá por onde começar? Saberá qual é o próximo passo?
- • O que o usuário vai tentar fazer a cada momento?
- O usuário perceberá que a ação correta está disponível?
 - • Onde está o elemento de interface correspondente ao próximo passo?
 - • Que ações a interface torna disponíveis?
- O usuário associará o elemento de interface correto à meta a ser atingida?
 - • O elemento de interface revela seu propósito e comportamento?
 - • O usuário consegue identificar os elementos de interface?
- Se a ação correta é tomada, o usuário perceberá que progrediu em direção à solução da tarefa?
 - • Como a interface apresenta o resultado de cada ação?
 - • O resultado apresentado tem correspondência com o objetivo do usuário?

Considerando-se cada um destes critérios, há algumas soluções típicas para as falhas mais comuns:

- “O usuário tentará atingir a meta correta?” Se a interface falhar neste ponto, há pelo menos três soluções possíveis:
 - eliminar a ação, combinando-a com outra ou deixando o sistema assumi-la;

- fornecer uma instrução ou indicação de que a ação precisa ser realizada;
 - alguma parte da tarefa pode ser modificada para que o usuário entenda a necessidade desta ação.
-
- “O usuário perceberá que a ação correta está disponível?” Se o usuário possui os objetivos certos mas não sabe que a ação está disponível na interface, a solução pode ser:
 - atribuir um operador mais óbvio para a ação. Isto geralmente requer um item de menu ou instrução, em vez de combinação de teclas;
 - atribuir à ação um controle menos visível mas como parte de uma estrutura em que possa ser facilmente encontrado, como um submenu.

 - “O usuário associará o elemento de interface correto à meta a ser atingida?”

Desde de que o designer transmita de forma que o usuário reconheça símbolos do seu cotidiano assim ele se familiariza com espontaneidade. Para isso o designer tem que ter essas informações bem esclarecidas, e qual o objetivo do sistema e seu público-alvo.

3. Métodos Experimentais

3.1 Métodos de Avaliação Empíricos

Segundo JORDAN (1998), a maioria dos métodos para avaliação de interfaces envolve a participação dos usuários. Alguns métodos são conhecidos com empíricos. JORDAN (1998) reforça que não pode haver a substituição do usuário na avaliação por um determinado produto. Sendo que ao longo do desenvolvimento do sistema tenha tido uma preocupação na ergonomia aplicada por seus responsáveis, pode acontecer na utilização do sistema o usuário deparar com algum esforço não esperado. Neste momento, os métodos que envolvem participantes terão um valor adicional, promovendo a descoberta de problemas de usabilidade até então não conhecidas. Normalmente, os usuários podem estar aptos para lidar com facilidade com certos sistemas que, de acordo com os princípios e convenções da ergonomia, esperava-se certo esforço durante a sua utilização. Vale lembrar que estes fatos só podem ser diagnosticados através do envolvimento de participantes (usuários em potencial de determinada interface) durante a avaliação da usabilidade.

O avaliador tem que ter o pleno conhecimento minucioso sobre as condições do teste. Certificar-se que todos testes que serão aplicados tenha as mesmas condições para todos os usuários, e de que o teste sendo executado permite a avaliação dos critérios desejados.

Podendo assim na preparação do teste o avaliador determinar o objetivo da avaliação e, em função deste, os critérios relevantes e pontos críticos, selecionar as

tarefas, selecionar usuários participantes, gerar o material para o teste e executar o teste-piloto.

Objetivo da avaliação em laboratório

O objetivo da avaliação pode ser tão amplo quanto verificar se o sistema é usável ou bem compreendido. Quando se prepara um teste o avaliador deve definir os critérios relevantes ou prioritários para alcançar o objeto desejado e pontos críticos da aplicação a serem avaliados. Normalmente os critérios prioritários são definidos durante na etapa de design, como sendo aspectos da qualidade de uso que se pretende que o sistema possua. Normalmente os pontos críticos do design estão relacionados nas decisões de design, para que as tarefas da utilização do sistema seja de uso freqüente.

Seleção de tarefas

Após os objetivos de a avaliação serem determinados, o avaliador tem que apontar quais as tarefas a serem executadas durante o teste aplicado, que mostrara indicadores sobre os objetivos. Na aplicação do teste de usabilidade, o avaliador definir também as medidas a serem observadas para cada aspecto que se deseja apreciar.

O avaliador depois de ter definidos as tarefas deve tomar cuidado com o tempo previsto para cada realização das tarefas. A execução de cada tarefa deveria levar

normalmente no máximo 20 minutos, e o teste deveria manter inferior a uma hora para não ficar cansativo para o participante (Preece et al., 2002).

Seleção de participantes

O processo de seleção de participante, o avaliador deve se preocupar com o perfil do participante. O objetivo da avaliação normalmente, é buscar um participante que tenha o perfil dos usuários que iram utilizar o sistema. O perfil do participante depende muito do tipo do sistema e qual seu público-alvo.

A seleção para os participantes geralmente se recomendam que seja equivalente de homens e mulheres, não que o sexo do participante seja uma das característica desejado. Tem que levar em consideração a experiências dos participantes com um sistema similar.

Além de definir quais usuários poderiam participar do teste, o avaliador deve definir quantos usuários farão o teste. Vale ressaltar que o objetivo destes testes não é chegar a resultados estatisticamente válidos, mas se ter indicações de como melhorar a qualidade de uso da interface. Tipicamente em testes com usuários se envolve de 5 a 12 usuários (Dumas e Redish, 1999). Nielsen por sua vez recomenda que 5 usuários participem da avaliação (Nielsen, 2000). Segundo ele, estudos mostram que este número apresenta a melhor relação custo-benefício. Isto porque o teste com um usuário é capaz de identificar aproximadamente 30% dos problemas da aplicação. Cada novo usuário, encontra 30% de problemas, destes uma parte

representa novos problemas, enquanto a outra representa problemas encontrados pelos usuários anteriores. Desta forma, a cada novo teste se reduz o número de novos problemas, e se aumenta o número de problemas já encontrados. Segundo Nielsen com 5 usuários se encontra aproximadamente 85% dos problemas da aplicação e o benefício dos novos erros encontrados vale o custo do teste executado. A Figura 2 abaixo ilustra o custo e benefício dos testes em função dos números de problemas encontrados e do número de testes executados. Nos casos em que a aplicação se destina a usuários de perfis distintos, Nielsen recomenda que para cada perfil identificado se faça a avaliação com 3 (três) usuários. Isto porque muitas vezes usuários de perfis distintos identificam os mesmos problemas.

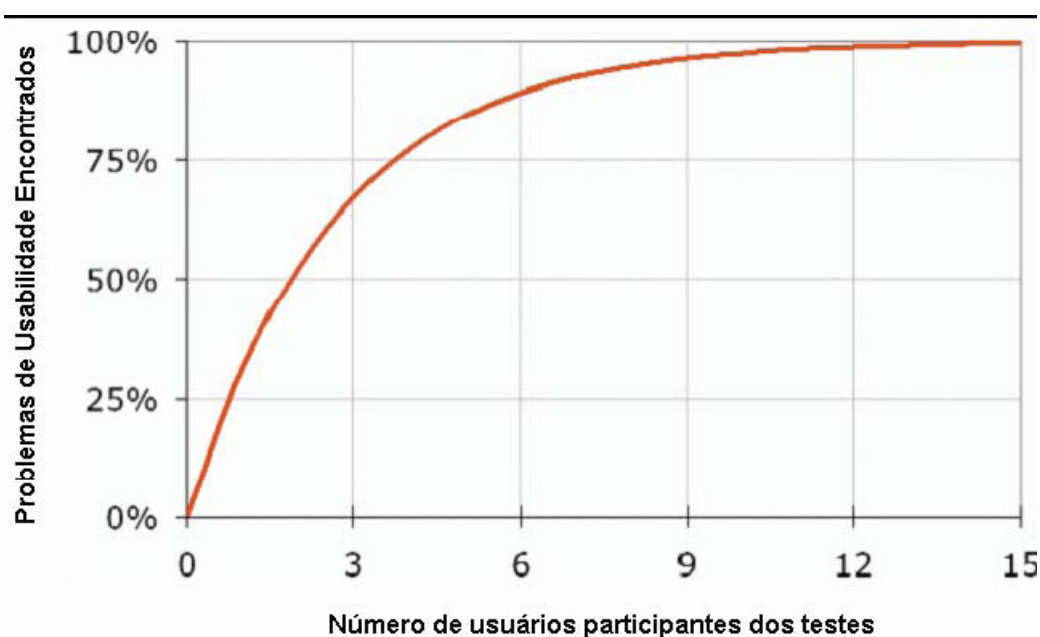


Figura 2 - Custo x benefício de execução de testes.
(Fonte: Nielsen, J.: Test with 5 Users, Alertbox, 2000)

Nielsen ressalta ainda que, embora um teste com 15 usuários permita potencialmente que se encontre todos os problemas de uma aplicação (ver Figura 6), vale mais a pena fazer três testes com 5 usuários do que um com 15. Isto porque com um teste com 15 usuários, todos os problemas poderão ser encontrados, mas a solução a ser desenvolvida após o teste não será avaliada e pode conter novos problemas. Em contrapartida três testes com 5 usuários, permitirá que se encontre potencialmente 85% dos problemas da aplicação no primeiro teste. A solução desenvolvida será também avaliada, e 85% dos problemas desta nova solução, sejam eles inseridos pela nova solução, sejam eles já existentes na versão anterior, mas não detectados pelo teste anterior, serão descobertos. Assim, incrementalmente se caminha para uma melhor solução final.

3.2 Composição do material para o teste

O avaliador após ter definidos os objetivos, tarefas e perfis dos desejados participantes para o teste, produz o material a ser utilizado. O material de avaliação contém um questionário, scripts da apresentação e a explicação de como será o teste para o participantes, formulários de consentimentos dos participantes e texto de descrição da tarefa e roteiros de entrevista ou questionários.

É formulado um questionário para conhecer o perfil dos participantes, para que encontre o perfil desejado. Os scripts de apresentação é a explicação a todos os participantes de como será a aplicação do teste assim todos receberam as mesmas informações. Antes de dar início ao teste devem colher os dados pessoais dos

participantes para que isto seja um consentimento formal contendo a assinaturas dos mesmos.

Todas as tarefas serão apresentadas aos participantes por escrito, desta forma esclarece como será a aplicação do teste, de uma descrição de uma situação plausível e contextualizada de uso. O objetivo de mostrar o cenário é motivar os participantes a interagir com clareza. A aplicação do teste devera ser de forma padrão a todos os participantes contendo as mesmas informações.

Teste piloto

O teste piloto é um protótipo do teste, é nele que pode-se avaliar a qualidade do material do teste. O foco principal do teste piloto é verificar se os participantes conseguiram entender com clareza as informações passadas, se o tempo de execução do teste esta dentro do que foi previsto e se é viável. Pode-se aproveitar o teste piloto para praticar a habilidade do avaliador para deixar o participantes à vontade para o teste. Assim permite que os dados coletados durante o teste possam dar mais visão como um todo do teste.

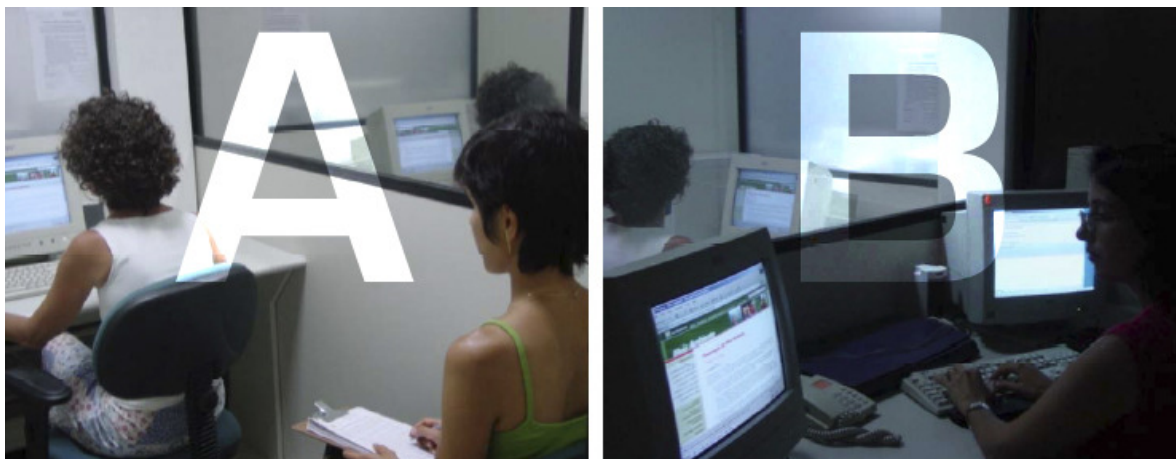
Recomenda-se que o teste piloto seja feito até que não tenha mais nenhuma alteração sobre o material a ser aplicado. Caso isto não seja possível, deve-se garantir no mínimo a execução de um teste piloto. A aplicação do teste piloto terá mais garantia sobre seu resultado ser os participantes tiver o perfil requisitado. Caso não tenha esse perfil de participante por motivo de custo elevado, ou acesso a eles limitados, pode-se executar com colegas.

3.3 Testes em Laboratórios

A escolha do laboratório deve ser um ambiente adequado para os testes, a ética é uma questão muito importante no que envolve o processo do teste com participação de outros seres humanos, e os participantes têm que se sentir à vontade, para que os participantes possam agir de forma naturalmente quanto consigam neste ambiente controlado e artificial.

Ambiente da aplicação do teste

Normalmente os teste são aplicados em laboratórios de testes com o usuário (Figura 7). Geralmente os laboratórios costumam ter duas salas onde, uma sala fica o usuário fazendo o teste (Figura 7 a) e a outra para observação do teste (Figura 7b). A divisão de uma sala para outro é por um vidro espelhando, de forma que o usuário não consegue perceber que tem alguém na outra sala. A sala onde o usuário realiza o teste é composta por um computador e espaço para um participante e um avaliador. A outra sala de observação costuma ter um monitor que reproduz o que o usuário esta testando, sem que o mesmo veja que tem outra pessoa observando-a, e um computador para anotações do teste. Também pode ter câmeras de vídeo e gravadores na sala onde esta sendo aplicado o teste ou na outra de observação, ou ambas uma em cada ambiente.



(A) participante e um avaliador na sala.
(B) avaliador observando por trás do vidro espelhado.

Figura 3 – Laboratório de comunicabilidade do Grupo de Pesquisa em Engenharia Semiótica (SERG), na PUC-Rio.

Quando não se tem um laboratório fixo como o citado acima, pode montar um laboratório móvel (com um computador, câmeras de vídeo, ou um sistema de registro de interação) e se utilizar em uma sala disponível no local do teste para servir como laboratório. Quando a avaliação pede em um local específico ou dificuldade dos participantes, pode ocorrer essa opção de laboratório móvel.

Éticas

Na execução de testes que envolvem outros seres humanos o avaliador deve estar atento às questões éticas envolvidas e a como lidar com elas. Alguns governos, como o dos EUA, regulamentam o processo de teste e têm exigências formais e burocráticas sobre sua execução. Para garantir a proteção aos usuários e o atendimento a exigências éticas as seguintes diretrizes foram propostas (Preece et al., 2002):

- Explicar aos participantes os objetivos do estudo sendo feito e exatamente como deverá ser a participação deles. Deve-se deixar claro o processo de teste, o tempo aproximado do teste, o tipo de dado sendo coletado e ainda como os dados serão analisados.
- Deixar claro as expectativas de anonimato dos usuários, deixando claro que dados particulares identificados durante o teste não serão divulgados.
- Certificar-se de que os usuários entendem que a qualquer momento podem interromper o teste, caso desejem.
- Sempre que trechos de depoimentos dos usuários forem utilizados eles devem ser anônimos e deve-se retirar descrições ou trechos que permitam a identificação do usuário. Nestes casos, deve-se requisitar autorização prévia do usuário e de preferência mostrar-lhe o relato a ser divulgado.

O participante antes de realizar o teste deve consentir por escrito que concorda e está ciente o motivo desse teste que irá participar, e neste documento deve conter a assinatura tanto dele e do avaliador. O formulário deve permitir que o participante em acrescentar novas condições ao acordo, caso se necessário.

Aplicação do teste

O avaliador faz a apresentação e a explicação do teste, de forma clara e objetiva sendo essa apresentação do teste para todos os participantes. Apesar de o uso de script reforçar o sentimento de ambiente artificial, deve se esforçar para que o usuário fique à vontade. Deve-se então pedir a ele que leia, preencha e assine o formulário de consentimento de execução do teste, que deve também ser assinado

pelo avaliador. Algumas vezes se utiliza questionários para se conhecer especificidades sobre o usuário. Isto pode ser feito antes ou após o teste.

Ao Termino do teste, costuma-se entrevistar o participante ou lhe pedir para preencher o questionário. Quando se tem uma entrevista pode-se ser tirar dúvidas sobre as ações observadas do usuário, ou sobre as motivações ou expectativas que o levaram a realizá-las. Além disso, através da entrevista ou questionário pode-se colher a opinião pessoal do usuário sobre sua experiência, sua satisfação com o sistema e sugestões.

Análise dos dados coletados

Note-se que durante o teste diferentes dados são coletados de diferentes formas: registro de uso, anotações de observação, preenchimento de questionários e condução de entrevistas. Na etapa de análise o avaliador deve analisar os dados coletados durante o teste para todos os usuários e a partir dele gerar o relatório do teste. O grande volume de dados normalmente implica um longo tempo necessário para sua análise. Em alguns casos, alguns dos dados coletados não são sempre analisados, mas apenas para verificação de comportamentos específicos ou para tirar dúvida sobre algum outro dado analisado. Por exemplo, quando se tem observadores fazendo anotações durante o teste, a gravação de vídeo pode ser utilizada apenas em pontos onde uma anotação feita não ficou clara, ou se gostaria de rever a reação do usuário durante um trecho da interação.

Normalmente, quanto mais cedo se consegue fazer a análise dos dados após sua coleta melhor, uma vez que o que se foi observado ou relatado está fresco na memória do avaliador.

O relatório costuma descrever os testes feitos, os problemas encontrados, e dependendo do nível de informação que o avaliador tenha sobre as intenções e decisões de design, pode conter também hipóteses sobre as causas dos problemas observados. O relatório não necessariamente inclui propostas de redesign para se solucionar os problemas identificados.

3.4 Testes de Usabilidade

O teste de usabilidade é executado em laboratório e tem por objetivo permitir que se apreciem os fatores que caracterizam a usabilidade de um software, ou seja, facilidade de aprendizado, facilidade de uso, eficiência de uso e produtividade, satisfação do usuário, flexibilidade, utilidade e segurança no uso (Nielsen, 1993; Preece et al., 2002).

Quais destes fatores devem ser priorizados deve ser definido no projeto de design. Através do teste procura-se quantificar o desempenho do usuário. Para isso durante a preparação de testes em laboratório descrita anteriormente, para cada medida a ser observada, deve-se definir quais são os limites mínimos aceitáveis, os máximos possíveis (em outras palavras, o melhor e pior caso) e também o valor almejado para a medida no projeto. A quantificação do desempenho normalmente envolve a medição do tempo e de ações de usuários. Apenas a satisfação do usuário se

distingue e normalmente é medida através da coleta de opinião do usuário e cujos limites mínimos, máximos e almejados costumam ser definidos em função da porcentagem de usuários que se dizem ou não satisfeitos com o software e o seu nível de satisfação. Alguns exemplos de medidas comumente utilizadas no teste de usabilidade são tempo gasto para se executar uma tarefa, número de erros executados, porcentagem de usuários a conseguirem se recuperar de um erro, ou porcentagem de usuários a se dizerem satisfeitos com a aplicação, ou a preferirem a aplicação a um outro sistema sendo utilizado.

Na análise dos dados coletados durante o teste de usabilidade o avaliador primeiramente classifica os problemas pela sua gravidade (Nielsen, 1998):

- Problema catastrófico: impede que o usuário termine sua tarefa
- Problema sério: atrapalha a execução da sua tarefa
- Problema cosmético: atrasa a execução e/ou irrita usuários

Além disso, para cada medida observada, ele verifica a distância para limites mínimos, máximos e almejados, determinando se o critério está em um patamar desejável ou não. Através dos dados o avaliador verifica o cumprimento de metas que tenham sido previamente definidas durante a etapa de design.

3.5 Testes de Comunicabilidade

Testes de comunicabilidade, como os de usabilidade, devem ser executados em laboratório. No entanto o seu objetivo é avaliar a interface com relação à qualidade da comunicação do designer para os usuários. Para isto, este método simula a comunicação do usuário para o designer sobre a interface. Isto é feito através de um pequeno conjunto de expressões que o usuário potencialmente pode usar para se exprimir em uma situação onde acontece uma ruptura na sua comunicação com o sistema (de Souza et al., 1999; Prates et al., 2000b).

No caso de testes de comunicabilidade, a gravação da interação do usuário com o sistema durante o teste deve ser feita, pois a análise será feita principalmente a partir deste registro. Além das anotações do observador durante o teste, as gravações em vídeo também podem ser feitas para enriquecer os dados, e permitir a verificação da reação do usuário relativa a algum trecho da interação observado.

A análise dos dados é dividida em 3 passos:

- Passo 1. Uma etiquetagem, que consiste em assistir às gravações da interação e atribuir a expressão apropriada nos momentos de ruptura da interação;
- Passo 2. Uma interpretação, que consiste em tabular e consolidar a informação obtida, ou seja, as expressões obtidas, associando-as a classificações de problemas de interação ou diretrizes de design; e

Passo 3. Um perfil semiótico, que consiste em interpretar a tabela resultante do passo anterior, dentro do quadro teórico da Engenharia Semiótica (de Souza, 1993; de Souza et al. 2001), em uma tentativa de se reconstruir a meta mensagem sendo transmitida pelo designer ao usuário através da interface.

Etiquetagem

Durante a etapa de etiquetagem, o avaliador assiste as gravações da interação feitas durante a coleta de dados. Ao observar uma ruptura da interação o avaliador associa à seqüência de ações problemática uma das expressões de comunicabilidade. O efeito desta etiquetagem é semelhante ao de um protocolo verbal (seção 3.7) “reconstruído” a partir das evidências de interação. A reconstrução é feita pelo avaliador, por meio de um conjunto de expressões, definido com o objetivo de ser o conjunto mínimo capaz de caracterizar suficientemente as rupturas de interação que acontecem durante o uso de uma aplicação. As expressões foram selecionadas com o intuito de serem naturais e espontâneas, e serem manifestações plausíveis por parte de usuários nestas situações. A seguir descrevemos o conjunto de expressões, seus significados e algumas ações de interface que caracterizam cada uma delas.

Cadê?

Ocorre quando o usuário sabe a operação que deseja executar mas não a encontra de imediato na interface. Um sintoma freqüente é abrir e fechar menus e submenus e passar com o cursor de mouse sobre botões, inspecionando diversos elementos de interface sem ativá-los.

E agora?

O usuário não sabe o que fazer e procura descobrir qual é o seu próximo passo. Os sintomas incluem vagar com o cursor do mouse sobre a tela e inspecionar os menus de forma aleatória ou seqüencial.

O que é isto?

Ocorre quando o usuário não sabe o que significa um elemento de interface. O principal sintoma consiste em deixar o cursor do mouse sobre o elemento por alguns instantes, esperando que uma dica seja apresentada.

Epa!

O usuário realizou uma ação indesejada e, percebendo imediatamente que isto ocorreu, desfaz a ação. Os sintomas incluem o acionamento imediato do *Undo* ou o cancelamento de um quadro de diálogo aberto indevidamente.

Onde estou?

O usuário efetua operações que são apropriadas para outros contextos, mas não para o contexto atual (por exemplo, tenta digitar um dado em um campo desabilitado; digita um comando em um campo de dado ou um dado no campo reservado para comandos). Um sintoma típico é desfazer a ação incorreta e mudar em seguida para o contexto desejado.

Assim não dá.

O usuário efetuou uma seqüência (longa) de operações antes de perceber que estava seguindo um caminho improdutivo. Os sintomas incluem o acionamento de *Undo* repetidas vezes ou o cancelamento de um ou mais quadros de diálogos abertos indevidamente.

Por que não funciona?

A operação efetuada não produz o resultado esperado, mas o usuário não entende ou não se conforma com o fato. O sintoma típico consiste em o usuário repetir a ação.

Ué, o que houve?

O usuário não percebe ou não entende a resposta dada pelo sistema para a sua ação (ou o sistema não dá resposta alguma). Os sintomas típicos incluem repetir a ação ou buscar uma forma alternativa de alcançar o resultado esperado.

Para mim está bom...

Ocorre quando o usuário acha equivocadamente que concluiu uma tarefa com sucesso. O sintoma típico é encerrar a tarefa e indicar na entrevista ou no questionário pós-teste que a tarefa foi realizada com sucesso. O observador, no entanto, sabe que se trata de um engano, provavelmente causado por uma falha de resposta do sistema ou modo de visualização inadequado para a tarefa atual.

Desisto.

O usuário não consegue fazer a tarefa e desiste. O sintoma é a interrupção prematura da tarefa. A causa pode ser falta de conhecimento, tempo, paciência, informação necessária, etc.

Vai de outro jeito.

O usuário não consegue realizar a tarefa da forma como o projetista gostaria que ele o fizesse, e resolve seguir outro caminho, geralmente mais longo ou complicado. Cabe ao avaliador determinar, se possível junto ao designer, qual é a forma preferencial de execução da tarefa.

Não, obrigado.

O usuário já conhece a solução preferencial do designer, mas opta explicitamente por uma outra forma de interação. O sintoma consiste em uma ocorrência da ação preferencial seguida de uma ou mais formas alternativas para se alcançar o mesmo resultado.

Socorro!

O usuário não consegue realizar sua tarefa através da exploração da interface. O sintoma é recorrer à documentação ou pedir explicação a outra pessoa.

Interpretação

Nesta etapa o avaliador deve associar as expressões identificadas a classificações de problemas de interação ou diretrizes de design. Utilizamos aqui uma classificação genérica que define os problemas de interação como sendo de navegação,

atribuição de significado, percepção, falha de execução da tarefa, e incompreensão ou recusa de *affordance*. Problemas de falha na execução da tarefa são os mais graves, uma vez que o usuário não consegue atingir o objetivo que o levou a usar a aplicação. Os de navegação se referem àqueles nos quais os usuários se “perdem” durante a interação com o sistema. Os de atribuição de significado, conforme o nome diz, acontecem quando o usuário não é capaz de atribuir um significado (relevante) a signos encontrados na interface. Os de percepção são quando os usuários não conseguem perceber alguma resposta do sistema ou seu estado corrente. Para a associação de expressões a problemas o avaliador poderia utilizar outras classificações de problemas, como por exemplo as 8 regras de ouro de Shneiderman (1998) ou ainda as heurísticas de Nielsen (1994).

A classificação proposta apresenta duas outras classes de problemas, incompreensão e recusa de *affordance*, que normalmente não fazem parte das classificações mais populares (e.g as citadas acima). No caso do problema de incompreensão de *affordance*, o usuário não consegue entender uma solução oferecida pelo designer, e acaba por executar a tarefa desejada de uma forma mais complicada, que não caracteriza a solução principal do designer. Finalmente, no caso de recusa de *affordance*, o usuário entende a solução principal oferecida, mas escolhe não utilizá-la e em seu lugar utilizar outra forma de interação que julga ser melhor. Esta outra forma pode ter sido prevista, pretendida, ou não, pelo designer.

A Tabela 1 (Prates e Barbosa, 2006) mostra como as expressões podem ser associadas a estas classes de problemas. Observe que para as expressões “Epa!”, “Onde estou?” e “Assim não dá” a associação a uma classe de problemas não é

unívoca e deve ser definida de acordo com o contexto em que foi observada a ruptura.

A partir da tabulação dos problemas encontrados o avaliador pode definir os pontos críticos da interação e gerar o relatório da avaliação.

Tabela 1 - Associação entre expressões e classes de problemas

Expressão de comunicabilidade	Problemas de Interação					
	Execução	Navegação	Atribuição de significado	Percepção	Incompreensão de <i>affordance</i>	Recusa de <i>affordance</i>
Cadê		X				
E agora?		X	X	X		
O que é isto?			X			
Epa!		X	X			
Onde estou?		X	X	X		
Assim não dá.		X	X	X		
Por que não funciona?			X	X		
Ué, o que houve?			X	X		
Para mim está bom...			X	X		
Não dá.	X					
Vai de outro jeito					X	
Não, obrigado.						X
Socorro!	X	X	X			

Perfil Semiótico

O perfil semiótico da aplicação deve ser feito por um especialista em Engenharia Semiótica (de Souza, 1993; de Souza et al. 2001). Segundo Prates e Barbosa, (2006) Neste passo o especialista interpreta a etiquetagem e tabulação feitas nos passos anteriores dentro do quadro teórico da Engenharia Semiótica, em uma

tentativa de se reconstruir a meta-mensagem sendo transmitida pelo designer ao usuário através da interface .

3.6 Testes de Usabilidade x Testes de Comunicabilidade

De acordo com Prates e Barbosa, (2006) maior diferença entre os testes de usabilidade e comunicabilidade está no conceito de qualidade de uso que eles pretendem apreciar, conforme seus próprios nomes indicam. Assim, testes de usabilidade pretendem avaliar a solução do designer, enquanto os de comunicabilidade buscam avaliar a comunicação sendo feita sobre esta solução. Os testes de usabilidade normalmente coletam dados quantitativos e buscam informar designers durante o ciclo de desenvolvimento quais critérios não correspondem ao objetivo almejado para o software. Testes de comunicabilidade, coletam dados qualitativos e têm por objetivo informar designers sobre pontos da sua solução que não estão sendo transmitidos com sucesso aos usuários.

3.7 Protocolos Verbais

Os métodos que envolvem protocolos verbais normalmente são testes executados em laboratório, nos quais os participantes explicitam aquilo que estão pensando, à medida que executam as tarefas propostas. A principal vantagem destes métodos sobre outros métodos de teste em laboratório é permitir que o avaliador tenha

acesso aos processos mentais dos participantes descritos por eles mesmos no decorrer da execução da tarefa. Desta forma avaliadores podem compreender melhor os comportamentos, sentimentos e atitudes dos usuários em relação à atividade sendo observada.

Um dos principais métodos de protocolo verbal utilizado é o de pensar alto (*think aloud*) (Ericsson & Simon, 1985). Neste método o participante executa os testes individualmente no laboratório e narra o que está pensando à medida que o faz. Se o participante fica em silêncio por um longo período de tempo, o avaliador pode lembrar o participante de continuar dizendo o que está pensando, com perguntas como: “O que você está pensando?” ou “O que você acabou de fazer?”. No entanto, esta interrupção pode atrapalhar a atividade do participante. A maior desvantagem deste método é que o participante faz duas coisas ao mesmo tempo: executa a tarefa, e narra suas ações e pensamentos.

Uma forma de amenizar as desvantagens do método de pensar alto é colocar dois participantes juntos para executar a tarefa. Desta forma, os participantes trocam idéias entre si sobre o que fazer, o que significa o que estão vivenciando e seus sentimentos e atitudes relativos à atividade e à aplicação. Ao fazerem isso eles explicitam seus pensamentos e os avaliadores passam a ter acesso ao que estão pensando. Este método normalmente é conhecido por método de co-descoberta.

3.8 Avaliação no Ambiente do Usuário

Uma outra forma de se fazer avaliações é ao invés de fazê-las em um ambiente controlado e artificial, fazê-las no ambiente natural dos usuários onde a aplicação será utilizada de fato. Enquanto testes no laboratório permitem a identificação de problemas de interface e interação, avaliações no contexto do usuário normalmente são utilizadas principalmente para se facilitar a introdução de tecnologia ou avaliar o uso sendo feito desta no contexto do usuário (Bly, 1997).

Em avaliações no ambiente do usuário, normalmente a coleta de dados é feita através da observação do uso sendo feito da aplicação e conversas com os usuários. Os dados podem ser armazenados como anotações, gravações de áudio ou vídeo, ou uma combinação de formas. Podem ser armazenados também artefatos ou quaisquer indicadores que auxiliem no entendimento de como as pessoas trabalham no seu próprio ambiente. Normalmente o dado coletado é qualitativo e a análise feita sobre ele interpretativa.

Existem duas abordagens principais sobre o papel do avaliador quando atuando no contexto do usuário (Preece et al., 2002). Na primeira abordagem, o avaliador atua apenas como **observador** e deve se esforçar para registrar os acontecimentos e **não interferir** no contexto. Na outra abordagem, o avaliador atua como **participante** do contexto e tem por objetivo explorar ao máximo os detalhes envolvidos no contexto e nas ações que sucedem nele. Neste caso abordagens etnográficas são apropriadas e bastante utilizadas.

Independente da abordagem de atuação do avaliador utilizada, com freqüência o avaliador utiliza-se de esquemas de apoio à sua observação. O objetivo destes esquemas é guiar a observação, e organizar os dados sendo coletados.

Um exemplo de esquema é o proposto por Robson (1993):

- Espaço: Como é a disposição física do ambiente e como ele está organizado?
- Atores: Quais são os nomes e detalhes relevantes das pessoas envolvidas?
- Atividades: O que os atores estão fazendo e por quê?
- Objetos: Que objetos físicos, como por exemplo móveis, estão presentes?
- Atos: O que cada um dos atores está fazendo?
- Eventos: O que está sendo observado é parte de algum evento?
- Objetivos: Que objetivos os atores estão tentando atingir?
- Sentimentos: Qual o humor do grupo e dos indivíduos?

Uma vez feita a avaliação no ambiente do usuário recomenda-se rever as anotações e registros tão logo quanto possível (de preferência antes de passadas 24hs) para que os dados estejam frescos na memória e para que se possa explorar detalhes e descartar ambigüidades com outros observadores ou com as pessoas sendo observadas. Além disso, na preparação para a avaliação, aspectos sociais e de relacionamento com as pessoas sendo observadas devem ser levados em consideração, tais como de que forma ganhar a confiança das pessoas envolvidas na observação e como lidar com questões delicadas.

4. Requisitos para montagem do laboratório

Para montagem do laboratório de avaliação de interfaces humano-computador, foi levado em relevância uma das referências o modelo sugerido por Jakob Nielsen em seu livro *Usability Engineering*, de 1993 (Figura 4)

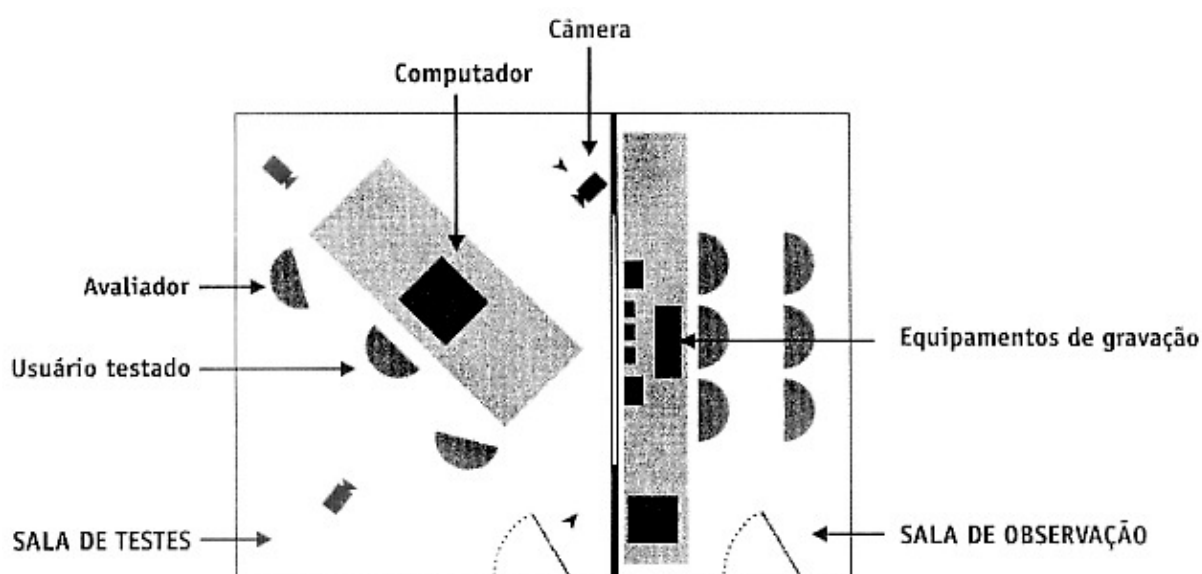


Figura 4 Planta sugerida por Nielsen.
(Fonte: Nielsen, J. (1993) *Usability Engineering*. Academic Press.)

Em cima da referência do modelo da (figura 4), foi estudado um rascunho do pré-projeto do laboratório (Figura 5) e um layout (Figura 6) pode ser visto mais claramente todos os requisitos para montagem desde das aparelhagem para que o teste seja registra das mais diferentes formas.

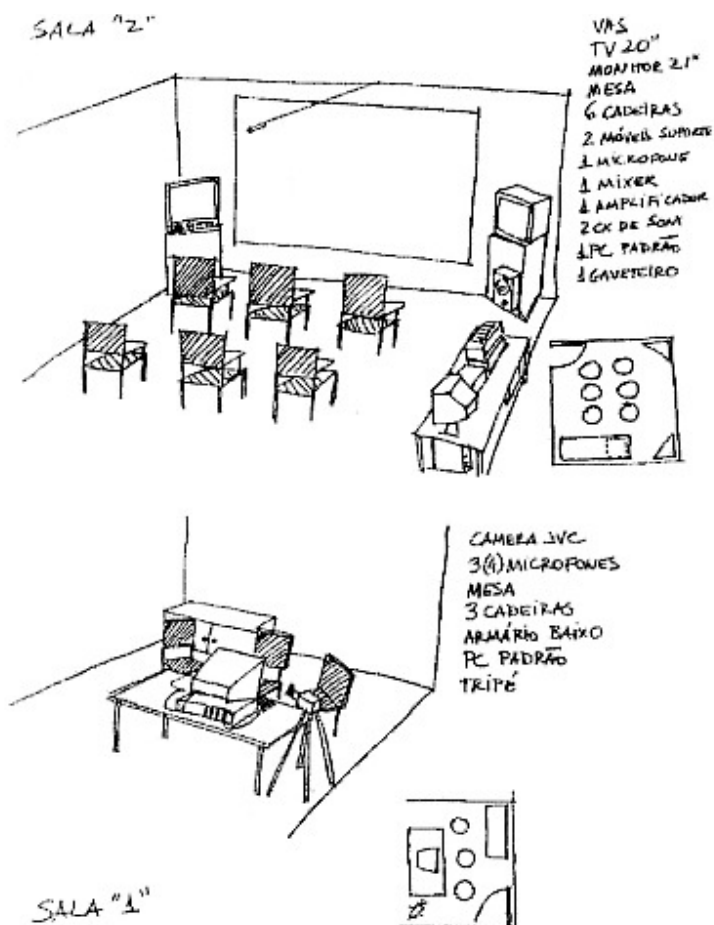


Figura 5 rascunho do pré-projeto.
 (Fonte: Cybis W,; Betiol A. H.; Faust R. (2007) "Ergonomia e Usabilidade Conhecimentos, Métodos e Aplicações")

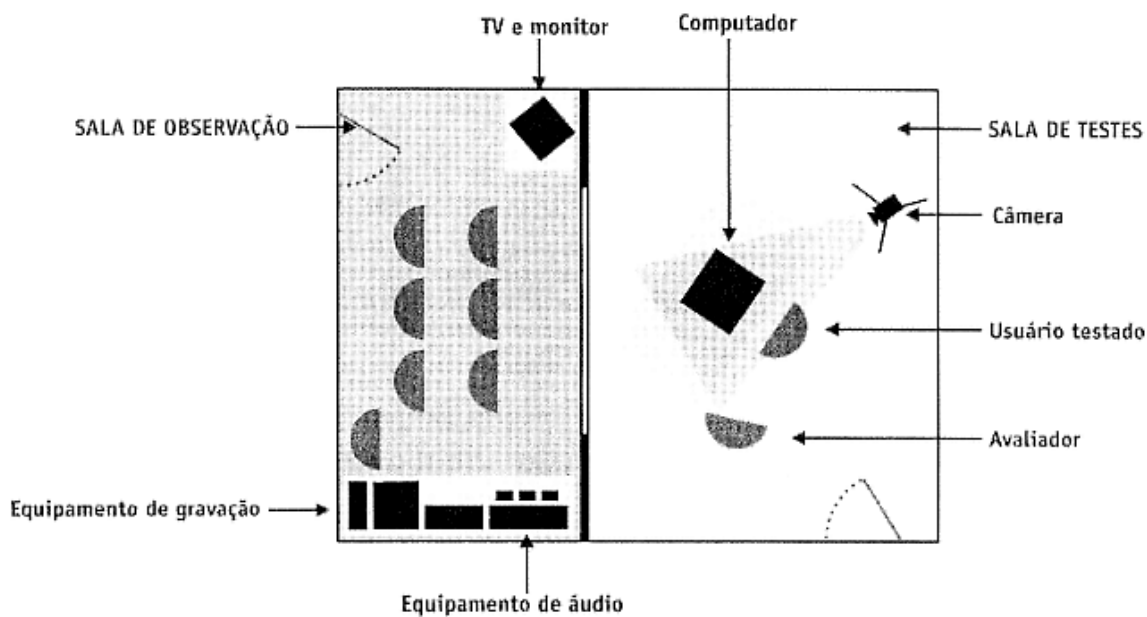


Figura 5 rascunho do pré-projeto.
 (Fonte: Cybis W,; Betiol A. H.; Faust R. (2007) "Ergonomia e Usabilidade Conhecimentos, Métodos e Aplicações")

5. Considerações finais

A usabilidade é uns dos fatores principais para o desenvolvimento de software e aplicações para WEB. Atualmente muitos estudos tem sido realizados com o objetivo de avaliar interfaces. Observa-se que existe a tentativa de levar estudos teóricos para o conforto empírico, ou seja, experimental. Neste sentido, o presente trabalho mostra como podem ser realizadas avaliações experimentais, culminando na especificação de requisitos físicos para a construção de um laboratório de avaliação de interfaces humano-computador.

Referências

- Prates, R.O e Barbosa, S.D.J. Disponível online em <http://www.dimap.ufrn.br/~jair/piu/artigos/avaliacao.pdf> [último acesso: novembro/2006]
- Asdi, K. and Daniels, H., (2000) “ 'Learnability' Testing in Learner-Centered Design”. In *CHI 2000 Extended Abstracts*.
- Barbosa, C. M. O., de Souza, C. S., Nicolaci-da-Costa, A. M., and Prates, R. O. P. (2002), “Using the Underlying Discourse Unveiling Method to Understand Organizations of Social Volunteers”, *Anais do V Simpósio sobre Fatores Humanos em Sistemas Computacionais (IHC 2002)*, Fortaleza, 15-26.
- CYBIS W.; BETIOL A. H.; FAUST R. (2007) “Ergonomia e Usabilidade Conhecimentos, Métodos e Aplicações. São Paulo, Novatec Editora, 2007.
- de Souza, C.S. (1993) “The Semiotic Engineering of User Interface Languages”. *International Journal of Man-Machine Studies*, Vol.39, 1993, pp.753-773.
- de Souza, C.S.; Prates, R.O.; and Barbosa, S. D. J. (1999) “A Method for Evaluating Software Communicability”. *Anais do II Workshop sobre Fatores Humanos em Sistemas Computacionais (IHC'1999)*. Campinas, Artigo 28.
- de Souza, C. S., Barbosa, S. D. J., Prates, R. O. (2001) “A Semiotic Engineering Approach to User Interface Design”. *Journal of Knowledge-Based Systems*, Vol.14, Issue 8, 2001, pp 461-465.
- Jordan, Patrick W. (1998). Disponível online em [último acesso: janeiro/2007] http://www.eduardobrandao.com.br/puc/mestrado/eduardo-brandao_capitulo-08.pdf
- MEDEIROS, Ethel Bauzer. *Medidas Psico&Lógicas: Introdução a psicometria*. Rio de Janeiro, Ediouro, 1999.
- Moran, T. (1981) “The Command Language Grammars: a representation for the user interface of interactive computer systems. Em *International Journal of Man-Machine Studies* 15:3-50, Academic Press.
- Mack, R. & Nielsen, J. (1994) *Usability Inspection Methods*. New York, NY: John Wiley & Sons.
- Nicolaci-da-Costa, A. M. (2001), “Gerando conhecimento sobre os homens, mulheres e crianças que usam computadores: algumas contribuições da psicologia clínica”, *Anais do IV Workshop sobre Fatores Humanos em Sistemas Computacionais (IHC'2001)*. Florianópolis, 120-131.

- Nielsen, J. (1993) *Usability Engineering*. Academic Press.
- Norman, D. (1988) *Psychology of Everyday Things*. BasicBooks. HarperCollins Publishers, 1988.
- PASQUALI, Luiz. *Psicometria Teoria dos testes na psicologia e na educação*. Petrópolis, Rio de Janeiro, Editora Vozes, 2003.
- Peirce, C.S. (1931-1958). *Collected Papers*. Edição brasileira: *Semiótica*. São Paulo, Ed. Perspectiva (coleção estudo, n.46), 1977.
- Pinelle, D. and Gutwin, C. (2002) "Groupware Walkthrough: Adding Context to Groupware Usability Evaluation". *Proceedings of the CHI 2002*.
- Prates, R. O.; de Souza, C. S; Carey, T.; Muller, M. J. (2001) "Evaluating beyond usability evaluation". Presented as SIG at CHI 2001, Seattle, USA. Available at (<http://www.serg.inf.puc-rio.br>)
- Prates, R.O.; de Souza, C.S. (2002) "Extensão do Teste de Comunicabilidade para Aplicações Multi-usuário". *Cadernos do IME*, Volume 13. 46-56, 2002.
- RAUPP, Magdala; REICHLE, Adriana. *Avaliação: Ferramenta para melhores projetos*. Santa Cruz do Sil. EDUNISC, 2003.
- ROSSON, Mary Beth; CARROLL, John M. *Usability Engineering. Scenario – Based Development of Human Computer Interaction*. San Diego, CA, 2002.
- Squires, D. e Preece, J. (1999). Predicting quality in educational software: "Evaluating for learning, usability and the synergy between them". *Interacting with Computers*. Vol. 11 (1999) 467–483