

O conteúdo do presente relatório é de única responsabilidade do(s) autor(es).
(The contents of this report are the sole responsibility of the author(s).)

Estudo Comparativo de Métodos para
Avaliação de Interfaces Homem-Computador

Sílvia Chan *Heloísa Vieira da Rocha*

Relatório Técnico IC-96-05

Setembro de 1996

Estudo Comparativo de Métodos para Avaliação de Interfaces Homem-Computador ¹

Silvio Chan ²

Heloisa Vieira da Rocha ³

Universidade de Campinas - UNICAMP
Instituto de Computação - IC
Caixa Postal 6176 - Cep 13081-970
Campinas - SP - Brasil
Fone: (019) 239 3115 / Fax: (019) 239 7470

setembro de 1996

Resumo

Neste relatório é apresentado um estudo comparativo de métodos de avaliação de interfaces homem-computador. O propósito deste estudo é verificar a aplicabilidade destes métodos, confrontando parâmetros tais como o perfil dos avaliadores, o envolvimento do usuário e desenvolvedore, restrições de tempo e material, escopo de avaliação, passos e duração da avaliação, adaptação a tipos específicos de problemas de utilizabilidade, e outros. Com este estudo pretendemos fornecer uma visão comparativa e classificatória dos métodos de avaliação para auxiliar organizadores de avaliação na escolha de métodos e no planejamento de uma melhor abordagem de avaliação.

¹ Este trabalho é apoiado pelo CNPq/Brasil e FAPESP.

² Aluno do programa de mestrado em Ciências da Computação do Instituto de Computação / UNICAMP.

³ Professora Doutora do Instituto de Computação / UNICAMP.

Estudo Comparativo de Métodos para Avaliação de Interfaces Homem-Computador

Silvio Chan

Heloisa Vieira da Rocha

Universidade de Campinas - UNICAMP
Instituto de Computação - IC
Caixa Postal 6176 - Cep 13081-970
Campinas - SP - Brasil
Fone: (019) 239 3115 / Fax: (019) 239 7470

setembro de 1996

1. Introdução

O crescente reconhecimento da importância da avaliação de interfaces homem-computador tem motivado o surgimento de diversos métodos para verificação da qualidade, eficiência e efetividade das interfaces. Entretanto, esta rápida expansão tornou a avaliação uma atividade geralmente informal, não estruturada e, até mesmo, desorganizada. Estas características muitas vezes prejudicam o próprio resultado da avaliação, levando os interessados na avaliação a optarem por métodos nem sempre adequados ao tipo de avaliação pretendida.

A avaliação de uma interface pressupõe a definição de elementos e conceitos básicos para a boa condução da atividade, que deve resultar na certificação de sua qualidade, principalmente, através da detecção de problemas na interface que vão afetar o usuário, a tarefa executada e toda a organização que os envolve. Tais definições estabelecem o que deve ser avaliado e como deve ocorrer a avaliação, e incluem, a definição do objeto de avaliação, a escolha dos avaliadores, o aspecto a ser avaliado, o escopo e o tipo de avaliação, e outros fatores que irão influenciar na escolha de determinado método de avaliação.

Os métodos de avaliação espalham-se num domínio extremamente amplo de aspectos, fatores e parâmetros, que podem ser considerados durante a atividade, incluindo toda a gama de restrições e questões envolvidas no ambiente de desenvolvimento e avaliação. Assim, torna-se necessário estabelecer uma classificação ou visão comparativa das características dos diversos métodos a fim de possibilitar a seleção, por parte do interessado na realização da avaliação, do método mais adequado ao que se pretende proceder.

Este trabalho corresponde a um estudo teórico comparativo de métodos de avaliação de interfaces homem-computador com o objetivo de auxiliar no planejamento e na condução da atividade de avaliação de interfaces. Os elementos e conceitos que devem ser definidos antes de realizar-se uma avaliação são tratados na seção 2. A seção 3 descreve as características dos métodos de avaliação mais conhecidos, bem como discorre sobre problemas de utilizabilidade e sua influência na determinação da efetividade da

interface. Na seção 4, tais métodos são classificados e organizados de modo a possibilitar a comparação entre os mesmos. Finalmente, a seção 5 conclui o trabalho.

2. Requisitos para Avaliação de Interfaces

Existem grandes possibilidades de variação na definição dos elementos e conceitos básicos da avaliação. Dois elementos essenciais podem ser identificados: o objeto de avaliação e o avaliador. O objeto de avaliação é o alvo da verificação de qualidade. Tanto pode ser a interface inteira, parte dela ou o conjunto formado por interface-usuário e a interação resultante, e dependendo do aspecto da avaliação, pode atingir também material como *helps* e manuais. Qualquer que seja o objeto de avaliação, este deve estar de alguma forma disponível aos avaliadores. Isto inclui desde simples especificações, modelos em papel, protótipos até a interface completamente implementada, havendo ligação direta com o momento em que ocorre a avaliação, ou seja, simultânea ao desenvolvimento, inserida no ciclo de desenvolvimento ou após o produto pronto.

A função de avaliador também pode ser exercida de várias formas e por diversos tipos de pessoas. O avaliador pode ser uma única pessoa ou um grupo de pessoas, trabalhando em conjunto ou individualmente. Da mesma forma, um avaliador pode ser um novato, um usuário típico, um desenvolvedor, um *designer* de interfaces ou um especialista em utilizabilidade, todos com níveis de conhecimento variando em um amplo espectro de áreas.

O aspecto da avaliação diz respeito ao enfoque dado quanto a eficiência, efetividade ou ambas na interface. Entende-se por eficiência a capacidade de se realizar o trabalho consumindo-se poucos recursos, isto é, economia de tempo, financeira, de material, entre outros. Efetividade corresponde a capacidade de realização de trabalho corretamente, permitindo que o usuário especifique exatamente como deseja sua execução e possibilitando que resultados sejam entregues de forma clara e de acordo com as especificações. Juntos, estes aspectos ainda incluem a facilidade de uso, a capacidade da interface em se adaptar ao usuário, a facilidade de interação, influenciando na satisfação do usuário.

A definição do escopo da avaliação envolve o estabelecimento de limites para a atividade em relação ao objeto de avaliação. Tais limites incluem a porção do objeto a ser avaliada, o que deve ser observado e quais aspectos a serem levados em consideração. Da mesma forma, o tipo de avaliação determina a natureza da avaliação em relação a como ela deve ocorrer. Segundo [Treu 1994], o tipo de avaliação distingue-a em:

- **Descritiva:** simples taxação da qualidade da interface.
- **Diagnóstico:** inspeção para determinar aspectos positivos e negativos na interface (méritos e problemas). Onde grande parte dos métodos de avaliação se enquadra.
- **Explanatória:** determinação dos motivos pelos quais decisões de *design* e implementação que afetam a funcionalidade, o comportamento e o *design* da interface foram tomadas.
- **Prescritiva:** verificação de formas de modificar o atual estado da interface, quer melhorando, quer degradando determinados aspectos. Apenas a título de exemplo, melhoria de aspectos da interface pode ser a introdução de uma metáfora mais apropriada para um sistema, enquanto degradação pode ser a substituição de um estilo de interação por outro menos flexível, porém mais conhecido e acessível aos usuários alvos/típicos do sistema.

Tendo definidos o objeto de avaliação, os avaliadores, o aspecto, o escopo e o tipo de avaliação, o próximo passo é definir quais os critérios a serem utilizados na avaliação. Tais critérios referem-se a como será apoiado o processo de avaliação, destacando-se formas de categorização das interfaces, princípios de utilizabilidade, *guidelines*, análise de tarefas, modelos mentais, etc. Os critérios têm relação direta com o método de avaliação, embora, geralmente, não haja impedimento para a utilização de outros critérios além dos especificados pelo método para contemplar questões particulares de determinada abordagem de avaliação. Da mesma forma, também existem critérios para obtenção e apresentação dos resultados da avaliação. Estes critérios dizem respeito às sessões de discussão (*brainstorm*), análise de dados coletados, forma de taxação da gravidade de problemas, consideração quanto a decisões de *design* e implementação e o fato da avaliação ser relativa ou absoluta, ou seja, se a interface será avaliada tendo-se ou não algum padrão de comparação (outra interface e/ou *guidelines*).

A forma de obtenção dos dados a serem utilizados na avaliação deve ser definida de modo a determinar a necessidade de elementos além do objeto de avaliação e dos avaliadores. Tais elementos incluem ferramentas automatizadas (*software*) para auxílio na avaliação, recursos materiais ou equipamentos para registro de dados (áudio e/ou vídeo), material impresso para guiar a avaliação e formulários para registro de dados observados durante a avaliação. Para tanto, deve-se considerar a possibilidade de utilizar dados como registro de sessões de interação em forma de gravação audiovisual, registro das ações do usuário em forma de *log*, entre outros.

A definição de todos os elementos e conceitos juntamente com a escolha de um ou mais métodos de avaliação, correspondem a uma abordagem de avaliação, que, bem selecionada, se enquadrará nos objetivos dos organizadores da avaliação.

3. Métodos de Avaliação

Entende-se por método de avaliação um conjunto de procedimentos que define de forma clara como realizar a avaliação. Vários são os métodos existentes para avaliação de interfaces. No entanto, a grande maioria consiste em métodos de inspeção para detectar problemas de utilizabilidade. O termo problema de utilizabilidade é usado de forma bastante genérica referindo-se a qualquer problema no objeto de avaliação que afete ou impeça a eficiência e efetividade do trabalho do usuário ou sua satisfação em realizá-lo. Os métodos de inspeção baseiam-se essencialmente no conhecimento e na experiência seja de avaliadores ou de *design guidelines* publicadas. Estes métodos podem variar de informais a bastante formais, mas têm como ponto comum o fato de poderem ser comparados a um processo de depuração, tal como na programação tradicional. Dentre os métodos, existe a categoria dos métodos heurísticos que inclui:

- Avaliação Heurística
- Avaliação baseada em *guidelines*
- Comparação entre Interfaces
- *Walkthrough* Cognitivo
- *Walkthrough* Pluralístico

Em outra categoria estão os métodos empíricos ou experimentais, que se baseiam em testes realizados com usuários quer sob condições controladas (laboratório) quer no mundo real. O método que se destaca nesta categoria é o Teste de Utilizabilidade

Além dos métodos citados acima, existem outras abordagens que muitas vezes não representam, exatamente, métodos de avaliação de interfaces, mas que podem ser utilizados para apoiar a avaliação e ratificar seus resultados. As versões de produtos com propósito de avaliação (versões beta) são um exemplo deste tipo de abordagem.

3.1. Problemas de Utilizabilidade

Métodos heurísticos e empíricos encaixam-se muito bem no tipo de avaliação de diagnóstico. Todos procuram verificar a qualidade de interfaces através da constatação de ausência ou presença de problemas de utilizabilidade, de suas características de ocorrência e gravidade. Os problemas detectados são, na maioria, infrações a princípios básicos de utilizabilidade, que podem ser considerados *guidelines* que deveriam ser seguidas em qualquer interface para garantia de eficiência e efetividade. São nove os princípios mais conhecidos [Nielsen e Molich 1989]:

1. Utilizar diálogo simples e natural
2. Falar na linguagem do usuário
3. Minimizar a carga de memória do usuário
4. Ser consistente
5. Fornecer *feedback*
6. Fornecer saídas claras e bem indicadas
7. Fornecer atalhos (*shortcuts*)
8. Utilizar boas mensagens de erro
9. Prevenir erros

Além destes princípios, [Karat et al. 1992] acrescentam outros cinco:

10. Fornecer um *layout* visual intuitivo
11. Fornecer bom *help*
12. Permitir customização pelo usuário
13. Minimizar o uso e efeito de modos
14. Suportar a continuidade de dispositivos de entrada

A ponderação dos problemas de utilizabilidade de uma interface tem dois aspectos, um quantitativo e outro qualitativo. O aspecto quantitativo refere-se ao número de problemas de utilizabilidade encontrados na interface. Uma interface que apresenta um grande número de problemas tende a impedir o trabalho efetivo do usuário. O outro aspecto refere-se à gravidade destes problemas. Entende-se por gravidade a taxação do problema de acordo com o impacto que ele exerce sobre o usuário impedindo a realização efetiva de tarefas através da interface. O critério de taxação da gravidade de um problema leva em conta o tipo de problema, a frequência com que ele ocorre, seu escopo, além observações próprias do avaliador.

O tipo do problema relaciona-se diretamente com o princípio de utilizabilidade desobedecido. A frequência do problema diz respeito ao número de vezes que é encontrado durante a avaliação, quantos avaliadores encontraram o mesmo problema e a sua recorrência. Entende-se por recorrência a capacidade e a quantidade de vezes que efeitos de um problema manifestam-se e afetam a interação. Um problema de baixa recorrência é aquele que interfere com a interação apenas na primeira vez em que é encontrado, por exemplo, um ícone cujo significado não é claro, mas que uma vez esclarecido deixa de prejudicar a tarefa.

Um problema de alta recorrência é aquele que constantemente revela seus efeitos sobre a interação, como por exemplo, a indisponibilidade de uma função não refletir na desativação de seu menu ou a alteração de modos sem alguma forma de aviso ao usuário. O escopo do problema refere-se a sua generalidade, ou seja, o problema interfere em várias regiões da interface, ou mesmo do sistema, ou é isolado numa região ou função. A propósito de exemplo, se a interface não informa se um documento está sendo impresso, tem-se um problema isolado. Se a interface não é consistente com o restante do sistema ou não fornece *help* sensível ao contexto, tem-se casos de problemas de escopo mais amplo. Observações do avaliador são consideradas em avaliações altamente apoiadas em experiência e conhecimento pessoal. Embora isto possa levar a um julgamento subjetivo do problema, e subsequentemente da interface, tal consideração não deve ser deixada de lado, especialmente no caso de avaliadores especialistas, pois em muitos casos referem-se a problemas já encontrados em outras avaliações ou experimentados pelo próprio avaliador tornando-o hábil a determinar seu impacto.

De posse dos resultados da avaliação, a equipe desenvolvedora, pode determinar quais os principais pontos da interface que devem ser atacados num *redesign* e quais podem ser relegados a segundo plano, para economia de recursos. Esta é a importância da taxação de gravidade, orientar o ciclo de vida e, no caso de avaliação após o produto pronto, revelar as maiores falhas do objeto de avaliação.

3.2. Descrição dos Métodos Heurísticos de Avaliação

Nesta seção são descritos os métodos heurísticos de avaliação de interfaces. Juntamente à descrição crítica dos métodos, tabelas organizadas em etapas são fornecidas procurando caracterizar a estrutura de cada um deles.

3.2.1. Avaliação Heurística

A avaliação heurística é um método em que os problemas de utilizabilidade são detectados através da inspeção em profundidade da interface pelo próprio avaliador. O elemento fundamental deste método é o avaliador, responsável por encontrar os problemas que ferem os princípios de utilizabilidade estabelecidos como parâmetros de avaliação. O avaliador, a priori, pode ser qualquer pessoa, envolvida ou não no desenvolvimento da interface. Isto inclui desde pessoas sem experiência em avaliação de interfaces ou áreas relacionadas até especialistas em utilizabilidade. Embora tais pessoas possam ter experiência e conhecimento acumulado extremamente variáveis, todas devem ao menos ter conhecimento de quais são os princípios de utilizabilidade nos quais se apoia a avaliação e quais seus significados.

No caso de avaliação com avaliadores não-especialistas em utilizabilidade, a sessão de avaliação deve iniciar-se com a explicação dos princípios aos avaliadores e/ou distribuição de folhas nas quais estes princípios são enumerados, definidos, ou mesmo, exemplificados. Obviamente, avaliadores com experiência em utilizabilidade dispensam tal introdução.

Com base nestes princípios, os avaliadores passam a percorrer a interface e descrevem em formulários os problemas nela encontrados. Nestes formulários devem constar o problema encontrado (descrição), seu tipo (princípio infringido), como foi descoberto (ações executadas que levaram à identificação do problema), classificação quanto a seu escopo e frequência, e taxação de gravidade. A avaliação pode ser feita

verificando, individualmente, cada um dos princípios de utilizabilidade em questão através de um roteiro ou ser livremente conduzida pelo avaliador, embora a primeira opção force uma certa estruturação no método, que é essencialmente pouco formal quanto a atuação dos avaliadores, especialmente quando estes são especialistas. Essa estruturação favorece a objetividade na atividade de avaliação.

Experimentos mostram que o perfil dos avaliadores tem grande impacto no desempenho deste método. Nielsen, em [Nielsen 1992], descreve um experimento realizado com três tipos de avaliadores: novatos, regulares e duplos. A categoria de novatos incluiu avaliadores sem experiência em utilizabilidade. Os regulares eram formados por pessoas consideradas especialistas em utilizabilidade. Os duplos correspondiam a avaliadores com experiência em utilizabilidade e no tipo particular de interface sendo avaliada. Entenda-se por especialistas em utilizabilidade quaisquer indivíduos com formação em área relacionada com utilizabilidade ou fatores humanos, ou pessoas que já tenham trabalhado durante longo período com questões de tais áreas. Os resultados do experimento revelam que os avaliadores duplos encontraram um número de problemas muito maior que os regulares, e estes uma quantidade superior aos novatos, considerando-se avaliações realizadas individualmente. O trabalho em conjunto dos avaliadores também contribui para o aumento do desempenho da avaliação. Quer trabalhando simultaneamente ou quer individualmente com os resultados reunidos no final, grupos distintos, cada um formado por avaliadores de uma das categorias, revelam que o método é efetivo desde que o número de avaliadores seja dimensionado de acordo com as expectativas da avaliação. Vide figura 1.

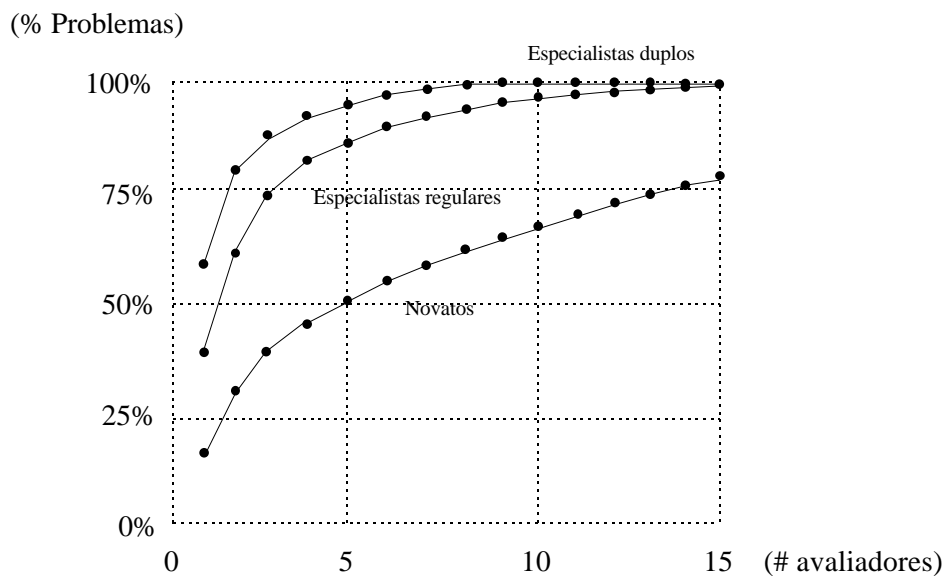


Figura 1: Variação do número de problemas detectados em relação tamanho do grupo de avaliadores. Extraído de [Nielsen 1992].

A superioridade dos avaliadores regulares e duplos frente aos novatos poderia indicar que este método de avaliação deve necessariamente ser conduzido por tais indivíduos, no entanto, especialistas em utilizabilidade são profissionais escassos e pouco disponíveis no mercado, e especialistas duplos, ainda mais raros. A escassez destes profissionais eleva o custo da avaliação, custo este que pode ser ainda maior se for necessário um grupo de cinco ou seis avaliadores especialistas, visto que, dificilmente, um único avaliador consegue detectar todos os problemas de utilizabilidade que seriam detectados por um grupo. Isto

leva a um impasse, pois o método de avaliação heurística foi proposto como um método de engenharia de desconto, ou seja, com objetivo de possibilitar engenharia com baixo custo. Assim, a utilização de grupos de especialistas em utilizabilidade ou especialistas duplos só se justifica se estes já estiverem disponíveis a um custo que possa ser suportado pelos interessados na avaliação ou se o objetivo da avaliação somente puder ser atingido através desta abordagem. Embora com desempenho menor que o de especialistas, avaliadores novatos também conseguem detectar um número considerável de problemas. Podem ser uma opção viável para proceder a avaliação, necessitando, obviamente, uma maior coordenação de pessoal para obter um bom resultado, pois o número de avaliadores novatos deve ser maior que o de especialistas para alcançar o mesmo desempenho de avaliação.

Outro experimento em [Jeffries et al. 1991], comparando quatro métodos de avaliação de interfaces, mostra a avaliação heurística como o método mais eficiente para detecção de problemas. No entanto, o método usou vários especialistas em utilizabilidade que trabalharam individualmente, e cujos resultados obtidos foram reunidos para compor o resultado final do método. Ao fim do experimento, a avaliação heurística foi reconhecida como o método que identificou o maior número de problemas, o maior número de problemas de gravidade séria e de mais baixo custo, considerando-se o total de problemas encontrados. Isto novamente favorece a avaliação heurística com especialistas, mas o mesmo experimento revela que este método identificou, também, o maior número de problemas menos sérios, que muitas vezes afetam apenas questões de eficiência e não a efetividade das tarefas, ou que são simples questões de gosto do avaliador, que portanto, podem ser muitas vezes encarados como julgamento individual.

1. Definição dos requisitos da avaliação: objeto, avaliadores, objetivos, escopo, aspecto, recursos necessários, etc.
2. Introdução: apresentação de informação aos avaliadores, incluindo objetivos, princípios e material de apoio (formulários, exemplos, manuais, etc.).
3. Avaliação da Interface: avaliadores testam a interface em busca de problemas de utilizabilidade. Os problemas encontrados devem ser registrados.
4. Discussão: avaliadores e outros indivíduos envolvidos na avaliação reúnem-se para discutir os problemas detectados e atribuir taxa de gravidade aos mesmos.
5. Apresentação dos resultados: divulgação dos problemas e determinação dos mais graves, que devem ser atacados num <i>redesign</i> .

Tabela 1: Etapas do método de avaliação heurística.

A tabela 1 resume as etapas da avaliação heurística, sugerindo um esquema simplificado para sua organização. Embora represente um custo alto, realizar uma avaliação heurística sem especialistas em utilizabilidade é impossível. Ao menos um indivíduo é necessário para apresentar os princípios aos não-

especialistas e realizar discussão para determinar a gravidade de problemas. Tais discussões ocorrem após as sessões de interação e correspondem a exposição de problemas encontrados e atribuição consensual de taxa de gravidade aos mesmos. Assim, neste caso, o especialista não é diretamente um avaliador, mas um orientador dos avaliadores não-especialistas.

Independentemente do perfil e experiência dos avaliadores escolhidos, este método exige que a interface esteja funcionalmente disponível, ou seja, implementada, ao menos parcialmente ou como protótipo para que os avaliadores possam utilizá-la. Isto limita a aplicabilidade do método no início de um ciclo de desenvolvimento de interface, retardando a descoberta de problemas.

3.2.2. Avaliação baseada em *Guidelines*

Guidelines correspondem a recomendações publicadas que tratam de questões de *design* e estilo em interfaces, podendo ser usadas como substitutas do conhecimento e participação de especialistas numa avaliação. Diversas corporações produtoras de *software* dispõem de suas próprias *guidelines* internas, por exemplo [Smith e Mosier 1986], que contribuem para consistência entre seus vários produtos e estabelecem regras ou princípios que, por experiência, melhoram a efetividade e a eficiência de interfaces. *Guidelines* podem ser consideradas conhecimento e experiência publicadas, e assim sendo, a avaliação baseada em *guidelines* apresenta certa faceta heurística. Da mesma forma, este é um método de inspeção baseado em princípios, pois os princípios de utilizabilidade nada mais são que *guidelines* de aplicação geral.

A avaliação de interfaces baseada em *guidelines* equívale a um *check-list* das propriedades, características, estilo e comportamento da interface frente ao que é recomendado pelas *guidelines*. A estrutura de *check-list* evidencia uma outra propriedade deste método de avaliação, ele pode ser considerado determinístico. Isto pode parecer antagônico em relação a definição do método como heurístico, mas estas propriedades se aplicam a componentes distintos do método. Seu aspecto heurístico deve-se às *guidelines*, derivadas diretas de experiência. O aspecto determinístico tem relação aos resultados, que na verificação estrita de *guidelines*, têm cobertura determinada por elas.

Este método de avaliação pode ser utilizado desde o início do desenvolvimento do produto e muitas vezes elimina a própria fase de avaliação, caso a interface seja desenvolvida rigidamente dentro das *guidelines* que seriam utilizadas na avaliação e pressupondo-se que a interface esteja obedecendo a todas elas. Tem-se neste caso um exemplo de avaliação de interface simultaneamente ao desenvolvimento da mesma. No caso de avaliação inserida no ciclo de desenvolvimento, *guidelines* podem ser aplicadas em modelos, especificações, protótipos e interfaces implementadas, possibilitando a cobertura de todo o ciclo.

A avaliação baseada em *guidelines* tem o menor custo dentre os métodos estudados, mas sua produtividade é restrita à abrangência das *guidelines* consideradas na avaliação. O baixo custo decorre da não necessidade de avaliadores especializados, pois a atividade pode ser conduzida pela própria equipe de desenvolvimento da interface - de fato é intenção do método utilizar desenvolvedores no papel de avaliadores. A abrangência restrita deve-se ao fato de que a maioria das *guidelines* referem-se a questões de âmbito geral entre interfaces e, raramente, abordam características pertinentes a um tipo específico de tarefa ou interface. Assim, como constatado em [Jeffries et al. 1991], este método é efetivo na detecção de problemas recorrentes e gerais (escopo amplo), mas falha na identificação de muitos problemas mais graves, que necessitariam de estudo mais aprofundado da interface, *guidelines* mais específicas, ou mesmo, participação de usuários e/ou especialistas.

Outro aspecto positivo deste método, além do baixo custo, é que avaliadores baseados em *guidelines* tendem a realizar avaliações mais completas que em outros métodos. A noção de avaliação completa refere-se a avaliação de todas as características e propriedades da interface, e não à detecção do maior número de problemas. A razão disto decorre do próprio sistema de *check-list* e das características das *guidelines*, que cobrem razoavelmente bem todas as propriedades das interfaces. Dessa forma, o resultado da avaliação não é tão influenciado pela forma como o avaliador percorre a interface como em alguns casos da avaliação heurística com especialistas. *Guidelines* portanto possibilitam um maior controle sobre os avaliadores.

Optar por este método significa estar sujeito à cobertura que as *guidelines* oferecem à avaliação. Formas de incrementar a efetividade do método incluem a adoção de avaliadores experientes de modo a superar a restrição das *guidelines* e compor uma abordagem misturando *guidelines* e avaliação heurística. Mesmo que avaliadores experientes não sejam utilizados, os resultados de uma avaliação através deste método mostram que dificilmente todos os problemas encontrados têm origem exclusivamente da avaliação baseada em *guidelines*, havendo sempre participação do conhecimento e experiência do avaliador, por menor que seja tal participação. No caso de uma avaliação simultânea ao desenvolvimento, deve-se esclarecer que ela não elimina a necessidade de avaliar a interface através de outros métodos para garantir que problemas graves não persistam na interface. As etapas deste método de avaliação são mostradas pela tabela 2.

1. Definição dos requisitos da avaliação: objeto, avaliadores, objetivos, escopo, aspecto, recursos necessários, etc.
2. Introdução: apresentação de informação aos avaliadores, incluindo objetivos, princípios, <i>guidelines</i> e material de apoio (formulários, exemplos, manuais, etc.).
3. Avaliação da interface: avaliadores fazem a conferência das propriedades da interface frente ao que é recomendado pelas <i>guidelines</i> . O problemas detectados são registrados em formulários.
4. Discussão: avaliadores e outros indivíduos envolvidos na avaliação reúnem-se para discutir os problemas detectados e atribuir taxa de gravidade aos mesmos.
5. Apresentação dos resultados: divulgação dos problemas e determinação dos mais graves, que devem ser atacados num <i>redesign</i> .

Tabela 2: Etapas do método de avaliação baseado em *guidelines*.

Avaliar com base em *guidelines* corresponde a uma boa alternativa em casos de escassez de recursos, avaliação no início do processo de desenvolvimento e para envolver os desenvolvedores na avaliação. Devido a suas características, *guidelines* são particularmente interessantes para enriquecer abordagens de

avaliação, mas devem ser cuidadosamente ponderadas no caso de sua opção como único método a ser utilizado.

3.2.3. Comparação entre Interfaces

A comparação entre interfaces não é propriamente um método para avaliação. Trata-se de uma abordagem simplificada de avaliação que pode ser baseada em heurísticas ou em *guidelines*, onde as recomendações são substituídas por decisões e soluções de estilo, *design* e implementação adotadas por cada interface. Esta abordagem de avaliação relativa pode ser classificada em:

- **Comparação com interface referência:** o objeto de avaliação é comparado a uma interface referência. Neste caso, a interface referência torna-se uma espécie de *guideline* e suas características são utilizadas como base de comparação em relação ao objeto de avaliação.
- **Comparação mútua:** não existe uma interface de referência. Os objetos de avaliação são comparados entre si, apontando assim, méritos e defeitos de cada um deles em relação ao outro. Trata-se de um processo heurístico, onde cada característica de uma interface é comparada à característica correlata de outra.

Embora seja uma abordagem não formal, a tentativa de esquematizá-la é válida, sendo isto proposto na tabela 3.

1. Definição dos requisitos da avaliação: objeto, avaliadores, objetivos, escopo, aspecto, recursos necessários, etc. Nesta etapa também deve ser determinada a divisão dos elementos de cada interface que serão comparados entre si. Esta divisão deve ser apresentada aos avaliadores na etapa 2.
2. Introdução: apresentação de informação tais como objetivos e material de apoio (formulários, exemplos, manuais, etc.) aos avaliadores.
3. Avaliação da interface: avaliadores fazem a comparação entre interfaces de acordo com a divisão determinada na etapa 1. O problemas detectados são registrados em formulários.
4. Discussão: avaliadores e outros indivíduos envolvidos na avaliação reúnem-se para discutir os problemas detectados e atribuir taxa de gravidade aos mesmos.
5. Apresentação dos resultados: divulgação dos problemas e determinação dos mais graves, que devem ser atacados num <i>redesign</i> .

Tabela 3: Etapas da abordagem de comparação entre interfaces.

Numa comparação é recomendável que as interfaces envolvidas na avaliação sejam de alguma forma similares, ou seja, não há sentido em comparar uma interface de *help* a uma de planilha de cálculos, ou uma GUI a uma interface baseada em telefone, salvo se ambas realizam a mesma tarefa. A razão disto é simples, ao se comparar dois objetos, procura-se colocar frente a frente, tarefas executáveis, funções, características e propriedades que existam em ambos. Assim, uma avaliação pode ser dividida em várias comparações de elementos similares a ambas interfaces.

A efetividade desta abordagem é variável e relativa. Ao escolhê-la, o interessado na avaliação deve estar ciente que sua produtividade não vai estar apenas relacionada à comparação, mas também à experiência do avaliador, especialmente se está sendo feita uma comparação mútua. Além disso, os problemas detectados, muitas vezes, não podem ter uma taxação absoluta, mas apenas uma indicação de ser melhor ou pior em relação ao elemento comparado.

A utilização desta abordagem é direta para avaliar informalmente produtos concorrentes e opções entre possibilidades de implementação, *design* e estilo.

3.2.4. *Walkthrough* Cognitivo

O método de *Walkthrough* Cognitivo corresponde a uma tentativa de tornar efetiva a prática da teoria cognitiva na avaliação de interfaces homem-computador. Para tanto, o *Walkthrough* Cognitivo utiliza como avaliadores membros da equipe de desenvolvimento da própria interface, que analisam tarefas selecionadas de forma a cobrir as metas gerais dos usuários típicos da aplicação. Trata-se de um método que pode ser aplicado desde o início do ciclo de vida do produto.

O que caracteriza este método como cognitivo é a consideração da atividade cognitiva do usuário na utilização da interface. O método leva em conta as metas do usuário, as tarefas que ele deve concluir para atingir tais metas, o conjunto de ações que devem ser executadas para concluir a tarefa, a percepção, a interpretação e a avaliação do resultado dessas ações. Ou seja, o método avalia a facilidade com que o usuário transpõe o que [Norman 1986] define como os Golfos de Execução e Avaliação ao usar a interface.

O método foi criado inicialmente para avaliar interfaces simples, que se apoiam no conceito de aprendizado através da utilização. Exemplos deste tipo de interfaces são os sistemas de caixa bancário eletrônico e quiosques de informações de centros de compra e aeroportos. Neste tipo de interface, o usuário executa tarefas muito bem definidas e geralmente simples. Dessa forma, aplicar o *Walkthrough* Cognitivo representa um processo relativamente simples. No entanto, quando o objeto de avaliação é uma interface mais complexa surgem questões que podem ser determinantes no sucesso deste método. Tais questões, definidos em [Wharton et al. 1992], são:

- A seleção, a cobertura e a avaliação das tarefas
- O processo de realização do *Walkthrough* Cognitivo
- O requisito de conhecimento para os avaliadores
- A influência de grupos de avaliadores
- A interpretação dos resultados

O *Walkthrough* Cognitivo realiza a avaliação analisando tarefas de forma a focar as atividades cognitivas do usuário, suas metas e conhecimento necessário, enquanto este realiza tais tarefas. [Wharton et al. 1992] estabelece que interfaces comuns podem suportar uma grande quantidade (dezenas ou centenas) de tarefas e variantes dessas tarefas e a seleção de quais devem servir como base de avaliação da interface é um processo que deve ser cuidadosamente realizado. A seleção das tarefas deve levar em conta:

- **O grau de realidade e complexidade das tarefas:** O grau de realidade de uma tarefa a ser avaliada pelo *Walkthrough* Cognitivo refere-se à proximidade dela em relação ao que o usuário realmente faz - um exemplo de tarefa menos realista em uma interface seria considerar que determinada tarefa tem apenas um caminho de execução. Complexidade refere-se ao número e dificuldade das ações envolvidas na tarefa. Tarefas menos realistas costumam ser avaliadas de forma mais rápida, mas corre-se o risco de não considerar problemas potenciais. Tarefas simples podem ser de fácil avaliação servindo como introdução à avaliação de outras mais complexas.
- **A fronteira dessas tarefas:** Tarefas muitas vezes não ficam limitadas à interface do sistema que está sendo avaliado. Muitas vezes, tarefas utilizam funcionalidade de sistemas subjacentes sem que o usuário perceba, embora os efeitos disso sempre apareçam como *feedback* na interface sendo avaliada. A título de exemplo, considere uma interface de sistema para gerenciamento de arquivos distribuídos e a tarefa de mover um arquivo de um diretório a outro. Fazer considerações quanto a fronteira dessa tarefa nesse sistema significa levar em conta ou não o fato de que a tarefa irá utilizar funções e recursos de sistemas subjacentes à interface, dentre os quais o sistema operacional, a rede e o sistema de armazenamento. Quanto se avalia tarefas que extrapolam a fronteira do sistema a que pertence a interface, problemas nos sistemas subjacentes podem vir a refletir na interface.
- **A quantidade de tarefas considerada suficiente:** A avaliação de uma tarefa é um processo que pode levar horas tornando a atividade tediosa. Escolher uma quantidade reduzida de tarefas, primando por uma cobertura adequada, é a melhor forma de evitar que problemas possam ser esquecidos devido a longa duração da avaliação.
- **A influência de variantes das tarefas:** Uma tarefa pode ser executada através de caminhos diferentes. Determinar quais caminhos devem ser considerados na avaliação reflete diretamente na quantidade de tarefas que podem ser avaliadas.
- **A granulosidade da avaliação:** Toda tarefa é constituída por uma série de ações. Tais ações são consideradas uma a uma, na seqüência em que são necessárias para concluir a tarefa, durante a avaliação. A granulosidade determina a que nível de resolução tais ações serão avaliadas, isto é, desde toques em teclas até ações de nível mais alto, como o preenchimento de um campo.
- **A consideração de subtarefas idênticas:** Diversas tarefas podem ser subtarefas em comum entre si. Tais subtarefas podem ser consideradas apenas uma única vez ou varias vezes nas tarefas avaliadas dependendo do contexto em que elas aparecem.
- **O tratamento de alto nível das tarefas:** Como dito, a avaliação através de *Walkthrough* Cognitivo considera as ações para executar tarefas. Reunir tais considerações no nível da tarefa representa um processo importante para estimular uma visão mais abrangente da tarefa pelo avaliador.

O método de *Walkthrough* Cognitivo, assim como outros métodos já descritos, baseia-se no preenchimento de formulários durante a avaliação. Tais formulários conduzem a avaliação indicando o que deve ser analisado, aspectos a considerar, forma de análise e taxaçoão ou classificação. Pelo fato de que uma grande quantidade de ações devem ser analisadas em cada tarefa avaliada, diversas sugestões são feitas para reduzir o desgaste dos avaliadores. Dentre tais sugestões incluem-se:

- **Estimular discussão entre os membros do grupo de avaliadores** sobre os problemas encontrados. As discussões devem manter foco na avaliação para evitar dispersão da atenção dos avaliadores e consumo excessivo de tempo.
- **Promover a troca de papéis dentro do grupo.** Não fixar funções. Por exemplo, cada um dos membros pode ser incumbido da responsabilidade de anotar as observações do grupo, preencher os formulários ou conduzir discussões de forma rotativa.

A figura 2 mostra um formulário específico para este tipo de avaliação.

Para responder as questões envolvendo valores percentuais, use os valores 0, 5, 10, 25, 50, 75 e 100%.

1. META do usuário:
2. Descreva a AÇÃO que deve ser executada:
 - 3.1. Está claro que existe a necessidade da AÇÃO para alcançar a META? Sim/Não/ Por quê?
 - 3.2. Haveria alternativa para evitar esta AÇÃO? Sim/Não/ Por quê?
 - 3.3. Que percentual de usuários poderia esquecer de executar a AÇÃO?
4. A AÇÃO é claramente possível de ser realizada pelo usuário? Sim/Não/ Por quê?
 - 5.1. A ligação entre AÇÃO e identificador dela é direta? Sim/Não/ Por quê?
 - 5.2. Que percentual de usuários poderia cometer engano nessa associação?
6. Ao executar a AÇÃO, existe *feedback* claro? Sim/Não/ Por quê?
7. Que percentual de usuários executaria essa AÇÃO corretamente?
8. Se a ação encerra a tarefa, existe indicação clara disto? Sim/Não/ Por quê?

Figura 2: Exemplo de folha de formulário para *Walkthrough* Cognitivo

Em contraste com os demais métodos descritos até o momento, o *Walkthrough* Cognitivo, não é diretamente um método de inspeção de utilizabilidade. Ele se utiliza da teoria cognitiva para encontrar dificuldades e/ou deficiências na conclusão de tarefas. De tais dificuldades é que se infere os problemas de utilizabilidade. Portanto, numa abordagem utilizando este método deve existir uma fase particular que é responsável pela interpretação das informações colhidas durante a avaliação.

Num *Walkthrough* Cognitivo, freqüentemente são utilizados termos de Ciência Cognitiva (exemplo: metas, tarefas e ações) nas sessões de avaliação e, especialmente, nos formulários. Como existe a pretensão de que os avaliadores sejam pessoas da equipe de desenvolvimento, não se pode obrigar os mesmos a terem conhecimento de tais termos ou mesmo realizar a avaliação rigidamente dentro do que poderia ser considerada uma avaliação cognitiva de tarefas. A maneira mais eficiente de se ter tal conhecimento corresponde a introdução de um especialista em cognição humana na liderança ou mediação do grupo. Dessa forma, os próprios desenvolvedores que participam do grupo passam a conviver com os conceitos e termos da área, que são apresentados e definidos pelo líder, tornando-se mais adaptados e preparados a futuras avaliações.

O número de membros num grupo de avaliadores pode influenciar no resultado da avaliação. Embora haja um consenso de que num grupo existe colaboração entre seus integrantes de forma a incrementar a produtividade, constatou-se em experimentos [Wharton et al. 1992] que grupos com mais de três membros tendem a prolongar demasiadamente as sessões de avaliação, especialmente devido a perda de foco nas discussões, dificuldade na divisão de responsabilidades e compartilhamento de recursos e informações, e queda da quantidade de problemas encontrados por avaliador.

A interpretação dos resultados obtidos em uma avaliação é uma questão geralmente tratada *ad hoc* numa avaliação baseada neste método. Formalmente, o *Walkthrough* Cognitivo não estabelece formas nem critérios para interpretar esses resultados. Existem indícios que mostram que este método parece trabalhar melhor com questões lingüísticas ou textuais [Wharton et al. 1992], como por exemplo botões ou itens de menu mal rotulados. Questões de natureza gráfica provavelmente possam ser melhor trabalhadas com auxílio de conhecimento em Engenharia Semiótica, embora não se tenha conhecimento de estudos unindo este método de avaliação de interfaces à Semiótica.

A tabela 4 mostra um esquema para realização de *Walkthrough* Cognitivo. Note que a fase 4 (Interpretação de Resultados) corresponde à fase anteriormente referida que procura, a partir das dificuldades do usuário, inferir problemas relacionados aos princípios básicos de utilizabilidade.

1. Definição dos requisitos da avaliação: objeto, avaliadores, objetivos, escopo, aspecto, recursos necessários, etc. Nesta fase devem também ser definidas as tarefas a serem avaliadas pelo método.
2. Introdução: apresentação de informações tais como objetivos, tarefas a serem avaliadas e material de apoio (formulários, exemplos, manuais, etc.) aos avaliadores. Nesta fase são atribuídos os papéis (funções) que serão exercidos por cada integrante do grupo de avaliadores (mediador ou líder, apresentador, avaliadores, registrador).
3. Avaliação da interface: avaliadores percorrem a interface de acordo com cada tarefa a ser avaliada, analisando cada uma das ações constituintes de cada tarefa. Os formulários conduzem esta fase, apresentada pelo apresentador. Os avaliadores anotam suas observações, preenchem os formulários e discutem eventuais pontos polêmicos levantados durante a fase. O mediador conduz o foco das discussões.
4. Interpretação de resultados: avaliadores e outros indivíduos envolvidos na avaliação reúnem-se para interpretar suas anotações e os formulários preenchidos. Desta interpretação surgirá a identificação dos problemas da interface e a atribuição da taxa de gravidade aos mesmos.
5. Apresentação dos resultados: divulgação dos problemas e determinação dos mais graves, que devem ser atacados num <i>redesign</i> .

Tabela 4: Etapas do método de *Walkthrough* Cognitivo.

As grandes vantagens deste método residem em:

- Poder ser aplicado já no início do ciclo de desenvolvimento.
- Permitir que analistas, *designers* e implementadores atuem como avaliadores. O *Walkthrough* Cognitivo pode ser visto como um laboratório para testes dos requisitos coletados por analistas e restrições técnicas de implementadores, contribuindo para diminuição do *gap* de comunicação entre estes profissionais e auxiliando a resolução de problemas em busca de melhores decisões de *design* e implementação.
- Corresponder a uma forma de reconhecer a importância de conhecer o usuário, a forma de como ele executa tarefas e realizar a análise cuidadosa de tarefas.

Suas desvantagens são:

- Não levar à identificação imediata de problemas de utilizabilidade.
- Poder induzir a soluções sub-otimais, ou mesmo errôneas. Por exemplo, propor a modificação de uma tela através da retirada de itens, quando tais itens são supérfluos apenas para as tarefas que foram avaliadas e são necessários para outras tarefas.
- No caso de não haver uma seleção criteriosa das tarefas, poder levar a uma visão muito isolada da interface e seus problemas.
- A utilização freqüente de conceitos de Ciência Cognitiva pode limitar a aplicação do método por desenvolvedores caso não haja uma preparação inicial que leve tais avaliadores a um domínio suficiente de termos e significados.
- Ser um método que, reconhecidamente, dispende tempo, como constatado em [Wharton et al. 1992] e [Jeffries et al. 1991].

Com intuito de reduzir o tempo gasto para avaliar uma interface através do *Walkthrough* Cognitivo, foi proposto uma variante chamada *Jogthrough* Cognitivo [Rowley e Rhoades 1992]. Esta variante apoia-se na automatização dos processos de registro de informações e observações levantadas tanto nas discussões quanto na avaliação em si, principal gargalo do *Walkthrough* Cognitivo, permitindo que a avaliação adquira maior rapidez. Esta automatização engloba o uso de software para agilizar o preenchimento dos formulários, e câmara de vídeo e computador para registro de eventos (*log*) sincronizados com o objetivo de capturar todos os momentos e fatos julgados relevantes que venham a emergir da avaliação ou de discussões. O *Jogthrough* Cognitivo permite que os avaliadores discutam os problemas com mais liberdade, não há rigidez no foco de discussão, estimulando a proposta de idéias e soluções durante a sessão de avaliação. Bem aplicada, esta variante permite agilizar todo o ciclo de desenvolvimento, dada a rapidez com que são obtidos os resultados da avaliação e o incentivo a proposta de soluções aos problemas encontrados logo durante a avaliação.

3.2.5. *Walkthrough* Pluralístico

O *Walkthrough* Pluralístico é um método de avaliação sistemático que reúne em um grupo de avaliadores, usuários finais, *designers*, desenvolvedores e especialistas em fatores humanos. A avaliação é baseada em cenários e protótipos de baixa-fidelidade (lo-fi). Protótipos de baixa-fidelidade [Retting 1994] são construídos em papel ou acetato, com várias peças cada uma representando elementos da interface (janelas, menus, botões, caixas, ícones, etc.) e suas possíveis variações. Este tipo de protótipo corresponde a uma opção aos protótipos convencionais de software, ditos de alta-fidelidade (hi-fi), e tentam maximizar o

número de avaliações dentro do ciclo de desenvolvimento reduzindo o tempo consumido pelas sessões. Dentre vantagens dos protótipos de baixa-fidelidade, pode-se citar:

- São de fácil e rápida construção e modificação.
- Eliminam a resistência dos desenvolvedores em relação a introdução de modificações na interface.
- Evidenciam que a avaliação está sendo feita num objeto em desenvolvimento e não num produto quase pronto.
- Eliminam a possibilidade de falhas de software que podem comprometer a avaliação.

A tabela 5 expõe um esquema para uma sessão de *Walkthrough* Pluralístico.

1. Definição dos requisitos da avaliação: objeto, avaliadores, objetivos, escopo, aspecto, recursos necessários, etc. Nesta fase devem também ser definidas as tarefas a serem avaliadas pelo método, por ele ser baseado em cenários.
2. Introdução: apresentação de informações tais como objetivos, tarefas a serem avaliadas e material de apoio (formulários, exemplos, manuais, etc.) aos avaliadores. Nesta fase são atribuídos os papéis (funções) que serão exercidos por cada integrante do grupo de avaliadores (introdutor, mediador, computador e observadores).
3. Avaliação da interface: o mediador apresenta cada ponto da interface a ser avaliado. Os observadores tentam utilizar a interface para executar as tarefas determinadas na etapa 1 e o computador se encarrega de representar a funcionalidade e comportamento da interface. Cada um dos observadores anota, individualmente, suas observações e os problemas que encontrou. O mediador estimula discussões e encoraja proposta de alterações que podem melhorar a interface.
4. Interpretação de resultados: avaliadores e outros indivíduos envolvidos na avaliação reúnem-se para interpretar as anotações. Desta interpretação surgirá a identificação de todos os problemas encontrados durante a sessão de avaliação da interface e a atribuição da taxa de gravidade aos mesmos.
5. Apresentação dos resultados: divulgação dos problemas e determinação dos mais graves, que devem ser atacados num <i>redesign</i> .

Tabela 5: Etapas do método de *Walkthrough* Pluralístico.

Durante uma sessão de avaliação, são atribuídos papéis aos membros da equipe de avaliadores. Tais papéis são divididos em introdutor, mediador, computador e observadores. As sessões de avaliação funcionam como uma exposição da interface, controlada pelo integrante que exerce a função de computador de forma

“teatral”, aos avaliadores, sobre a qual serão executadas tarefas em cenários pré-escolhidos. O mediador é o responsável por conduzir a sessão, enfocando cada questão a ser avaliada, dando instruções e estimulando discussão. Os observadores, dentre os quais podem estar usuários finais, atuam anotando em folhas individuais suas impressões e problemas encontrados em relação a cada questão enfocada pelo mediador. Como recurso de apoio, podem ser utilizadas câmaras de vídeo para acompanhar a avaliação e, principalmente, a atuação do computador sobre a interface, registrando todo o comportamento da interface.

A efetividade deste método está altamente relacionada ao ciclo *design-avaliação-redesign*. Quanto maior o número de iterações nesse ciclo, melhor poderá ser refinada a interface, mais problemas poderão ser identificados e mais propostas de soluções poderão ser experimentadas. Os méritos deste método são retrato das vantagens oferecidas pelos protótipos de baixa-fidelidade, sendo análogo em relação aos seus pontos negativos. Dentre as desvantagens do método, pode-se citar:

- Torna difícil para avaliar o “*look-and-feel*” de uma interface.
- Protótipos de baixa fidelidade não retratam detalhes.
- Limita o teste de modificações em sistemas reais, já existentes.

Apesar destas desvantagens, o método de *Walkthrough* Pluralístico revela-se útil, especialmente, no início do desenvolvimento. Encaminhar o processo de *design* e construção da interface com baixo custo e rapidamente, de forma que decisões iniciais básicas de qualquer processo de desenvolvimento não levem a problemas futuros. E, além disso, possibilita-se ter noção do que será obtido nas fases mais adiantadas do desenvolvimento. Este método evidencia o valor da atividade multidisciplinar no *design*, na solução de problemas em grupo e o *design* iterativo possível mesmo em restrições apertadas de tempo [Karat et al. 1992].

3.3. Teste de Utilizabilidade

Teste de Utilizabilidade é o nome genérico dado a alguns métodos empíricos para avaliação de interfaces [Dumas 1989]. Tais métodos diferenciam-se dos métodos heurísticos por empregarem controles experimentais [Karat et al. 1992] e têm como pontos em comum:

- A participação de sujeitos que representam uma amostra da população de usuários alvos da interface sendo avaliada nas sessões de avaliação.
- A utilização de cenários, tarefas típicas ou críticas.
- A coleta de dados originados da observação dos sujeitos que participam das sessões de avaliação.

O que distingue esses métodos são, basicamente, os procedimentos utilizados para envolver os sujeitos na avaliação e obter deles dados que levarão a inferir os problemas de utilizabilidade existentes na interface. Assim, pode-se citar os seguintes métodos mais conhecidos de Teste de Utilizabilidade:

- **Teste com pares de sujeitos:** este tipo de teste coloca pares de sujeitos frente a interface para a execução de tarefas. A utilização de pares tem por finalidade tomar vantagem do diálogo que se estabelece entre os sujeitos durante os testes, quando os sujeitos trocam informações e impressões sobre a interface, e tentam ajudar um ao outro na conclusão das tarefas..
- **Teste com sujeito e observador:** um observador (avaliador) é colocado “ao lado” do sujeito para acompanhar sua interação com a interface. A princípio, o avaliador apenas observa a progressão das

ações do sujeito para execução de tarefas, mas caso seja solicitado pelo sujeito e considere necessária, o avaliador pode prestar assistência.

- **Teste *Think-Aloud***: requisitar ao sujeito que verbalize seu raciocínio durante a sessão de teste da interface corresponde a outra maneira de observar suas impressões e detectar problemas de utilizabilidade. Estes três métodos podem se complementar, enriquecendo as informações colhidas pelo observador.

Embora os três métodos acima possam ser considerados como básicos em Teste de Utilizabilidade, existem laboratórios que utilizam estratégias bastante sofisticadas para capturar a maior quantidade possível de informações do usuário. Para tanto, além da observação do avaliador e da verbalização do raciocínio, são utilizadas câmaras de vídeo, gravadores de áudio e equipamentos para registro de eventos (*log*) para registrar todos os momentos da interação homem-computador. As informações registradas por esses dispositivos podem ser analisadas mais tarde pela equipe de avaliação e podem tirar dúvidas e fornecer subsídios valiosos para identificação de mais problemas de utilizabilidade.

Os controles experimentais empregados em Teste de Utilizabilidade contribuem para garantir o rigor da avaliação e fazem parte do planejamento ou definição dos requisitos da avaliação. Nesse planejamento são definidos os objetivos da avaliação, a decisão sobre qual método de Teste de Utilizabilidade adotar, a qualificação dos sujeitos, a quantificação da amostra de sujeitos da população de usuários, a seleção dos cenários, a priorização de tarefas, os tipos de métricas a utilizar (exemplo: tempo de conclusão de cada tarefa, número de requisições de assistência, tempo gasto para decisão, etc.), os recursos materiais necessários e a definição de todas as condições que podem afetar a avaliação e caracterizar um ambiente (exemplo: luminosidade, ruído, tempo, horário, temperatura, etc.).

A equipe de avaliação deve ter ao menos um especialista em interfaces que deve acompanhar explicitamente ou de forma remota (câmaras) cada sessão de avaliação. Envolver desenvolvedores tanto no planejamento quanto durante as sessões também é recomendável para que estes possam participar da definição inicial e visualizar as reações dos usuários frente à interface, permitindo-lhes importantes informações sobre o que alterar num *redesign*. Por ser um método que utiliza usuários alvos, existe a necessidade de uma abordagem que incentive a participação de tais sujeitos. A literatura, [Gomoll 1990] e [Dumas 1989], mostra que a relação equipe de avaliação-sujeitos deve evidenciar que numa sessão de avaliação o que está sendo testado é a utilizabilidade de um produto e não a capacidade do sujeito em utilizá-lo. Da mesma forma, deve-se tornar claro que o sujeito está na avaliação voluntariamente e que a qualquer momento, caso seja seu desejo, pode abandoná-la. A equipe de avaliação deve ressaltar a importância das informações que o usuário irá repassar a ela e incentivar a verbalização de seu raciocínio. Criar um ambiente mais próximo possível da realidade pode ser muito útil. O sujeito, um computador, uma mesa e uma sala reservada permitem simular a situação de um usuário que tenta utilizar a interface pela primeira vez. A prestação de assistência deve ser comedida durante a sessão. Tentar estimular a anotação por parte do usuário de questões que porventura representem dificuldades ou dúvidas para somente serem esclarecidas no final da sessão é uma boa estratégia.

Antes de iniciar a sessão de avaliação, o sujeito deve ser informado de qual a função do produto, as tarefas que deve tentar executar e sua ordem de execução. O usuário pode saber que vai ser observado durante a sessão, mas nunca esta observação deve ser entendida como uma forma de pressionar o sujeito. Provavelmente, isto aponta para a observação e comunicação remota entre observador e sujeito, permitindo que o último sinta-se mais a vontade durante a avaliação. As etapas de uma sessão de um Teste de Utilizabilidade são mostradas na tabela 6.

<p>1. Definição dos requisitos da avaliação: objeto, avaliadores, objetivos, escopo, aspecto, recursos necessários, etc. Planejamento englobando qualificação e quantificação dos sujeitos, definição dos cenários, condições ambientais, formas de observação, entre outros.</p>
<p>2. Introdução: apresentação de informações tais como objetivos, tarefas a serem avaliadas e material de apoio (formulários, exemplos, manuais, etc.) aos sujeitos. Nesta fase os sujeitos devem ser informados que tem total direito e liberdade de interromper ou abandonar a avaliação e esclarecidos quanto a sua importância e das suas anotações para o desenvolvimento do produto.</p>
<p>3. Avaliação da interface: de acordo com a seqüência de tarefas informadas ao sujeito na etapa de introdução, este passa a interagir com a interface buscando concluir tais tarefas. Durante essa interação, o sujeito deve ser observado das maneiras escolhidas no planejamento. As reações e impressões do sujeito devem ser registradas pois apontam diretamente a pontos positivos ou negativos da interface. O sujeito anota quaisquer questões que possam surgir durante a etapa para que a equipe de avaliação possa esclarecê-las no final.</p>
<p>4. Interpretação de resultados: as anotações do sujeito e os dados colhidos através da observação da sessão são interpretados. Desta interpretação surgirá a identificação de todos os problemas encontrados durante a sessão de avaliação da interface e a atribuição da taxa de gravidade aos mesmos.</p>
<p>5. Apresentação dos resultados: divulgação dos problemas e determinação dos mais graves, que devem ser atacados num <i>redesign</i>.</p>

Tabela 6: Etapas do método de Teste de Utilizabilidade.

Em relatórios de experimentos, [Karat et al. 1992] e [Jeffries et al. 1991], com Teste de Utilizabilidade, o método mostra-se efetivo. Em [Karat et al. 1992], o método revela-se com desempenho superior numa análise de custo-efetividade em relação aos métodos heurísticos. [Jeffries et al. 1991], que também compara o Teste de Utilizabilidade aos métodos heurísticos, mostra que o método detecta muitos problemas de gravidade séria e evita grande parte dos problemas menos sérios. No mesmo estudo, o método é classificado como o que apresenta o custo por problema detectado mais elevado. Esta contradição entre os dois estudos pode ser explicado pelas diferenças entre os procedimentos das abordagens adotadas em cada um deles [Karat et al. 1992]. Tais diferenças incluem a experiência e o número de avaliadores/observadores disponíveis, tamanho da amostra de sujeitos, tempo de duração das sessões de interação e o conjunto de tarefas utilizado, evidenciando que os custos numa avaliação não são absolutos.

O Teste de Utilizabilidade é capaz de encontrar problemas sérios que outros métodos dificilmente encontram. Por se apoiar em sujeitos, este método é mais propenso a detectar problemas do tipo “acidentais”, que seriam evitados, devido ao maior conhecimento, pelos especialistas e raramente são tratados em *guidelines*. [Jeffries et al. 1991] descreve um problema detectado por um sujeito ao apagar seu

homedir. Após tê-lo apagado, o sujeito não mais conseguiu “logar” no sistema sob avaliação. Naturalmente, especialistas dificilmente cometeriam a ato de remover seus próprios diretórios e *guidelines* simplesmente não costumam tratar estas questões. A característica de evitar problemas de menor gravidade do Teste de Utilizabilidade pode ser explicada pelo fato de que tanto sujeito (usuário) quanto observador atuam como “filtros” selecionando problemas mais graves. O usuário é capaz de absorver muitos dos problemas menos sérios por ser adaptável, e o observador por estar realmente interessado em problemas graves e não naqueles de caráter subjetivo ou cosmético.

Em relação ao custo do método e à sua aplicabilidade, deve-se ressaltar:

- O método não exige muito em relação a experiência dos avaliadores. Normalmente, um especialista em fatores humanos é suficiente para acompanhar a avaliação com um número reduzido de sujeitos.
- O método tende a consumir tempo. Além do tempo necessário para a sessão de avaliação, é necessário, em muitos casos, tempo para treinamento dos sujeitos. Além disso, o tempo gasto na atividade de avaliação é proporcional ao número de sujeitos envolvidos. Cada especialista em fatores humanos pode acompanhar apenas uma sessão por vez. Pode-se aumentar o número de especialistas para diminuir o tempo necessário para conclusão da avaliação, mas isto eleva o custo com avaliadores.
- Realizar um Teste de Utilizabilidade pressupõe uma interface em fase adiantada de desenvolvimento e com funcionalidade acoplada, ou um protótipo fiel ao produto final. Isto, de certa forma, restringe a aplicação do método em fases iniciais e prejudica o *redesign* pois numa fase adiantada, a introdução de modificações é tecnicamente mais difícil e a própria equipe de desenvolvimento torna-se mais relutante. Envolver desenvolvedores na avaliação pode ser então uma forma de amenizar esse efeito.
- Tentar envolver os usuários desde o princípio do desenvolvimento da interface pode ser útil. Assim, os usuários participam das decisões de *design* e acostumam-se melhor a interface que lhes é dirigida. Esta participação dos usuários pode contribuir na eliminação de grande quantidade de problemas nas fases iniciais e, conseqüentemente, reduzir custo e problemas num futuro Teste de Utilizabilidade.

Por ter um custo relativamente alto, quanto a tempo consumido e avaliadores necessários, tem-se buscado alternativas ao Teste de Utilizabilidade. A engenharia de desconto, que oferece a Avaliação Heurística, é uma das propostas [Jeffries e Desurvire 1992], mas o Teste de Utilizabilidade continua com vantagens que o tornam muito importante:

- Permite observar as reações e impressões de usuários aos quais a interface é dirigida.
- Não necessita de grande número de especialistas.
- Concentra-se em problemas realmente graves.
- Os problemas realmente são de impacto aos usuários.

4. Classificação dos Métodos de Avaliação

O objetivo desta seção é apresentar uma tentativa de classificação dos métodos de avaliação descritos na seção anterior. A tabela 7 classifica os métodos em função de aspectos que possam afetar sua facilidade de aplicação e adaptabilidade. A tabela 8 sugere uma adequação dos métodos a tipos de problemas específicos. Nessas tabelas estão expostos parâmetros que, usualmente, influem nas decisões e no planejamento de uma abordagem de avaliação, dentre os quais, requisitos, recomendações e custos. O conteúdo das tabelas reflete o que pode ser encontrado no decorrer do planejamento e execução da avaliação de interfaces e procura ser consistente ao que é observado em experimentos práticos.

Aspectos de Aplicabilidade e Adaptabilidade						
Métodos Aspectos	Avaliação Heurística	<i>Guidelines</i>	Comparação	<i>Walkthrough</i> Cognitivo	<i>Walkthrough</i> Pluralístico	Teste de Utilizabilidade
Estágio do Objeto	Protótipo/ Produto	Especificações/ Protótipo/ Produto	Produto	Especificações/ Protótipo/ Produto	Protótipo	Protótipo/ Produto
Momento do Ciclo	Intermediário/ Final	Inicial/ Intermediário/ Final	Final	Inicial/ Intermediário/ Final	Inicial/ Intermediário	Intermediário/ Final
Envolvimento de Usuários	Possível	Não	Não	Não	Sim	Sim
Envolvimento de Desenvolvedores	Não Recomendável	Sim	Sim	Sim	Sim	Possível
Necessidade de Conhecimento Adicional (1)	Não	Não	Não	Sim	Sim	Sim
Necessidade de Especialistas	Sim	Não	Sim (2)	Sim	Não	Sim
Papel dos Especialistas	Avaliadores	-	Avaliadores	Mediador/ Líder de Equipe	-	Avaliadores (Observadores)
Tipo de Especialista	Avaliação de Interfaces	-	Avaliação de Interfaces	Ciência Cognitiva	-	Avaliação de Interfaces
Consumo de Tempo	Baixo	Baixo	Baixo	Alto (3)	Baixo	Alto (4)
Necessidade de Equipamentos (5)	Não	Não	Não	Recomendável	Recomendável	Recomendável
Utilização de Cenários (6)	Recomendável	Não	Não	Sim	Sim	Recomendável
Deteção Direta de Problemas	Sim	Sim	Sim	Não (7)	Sim	Sim

(1) Conhecimentos adicionais aos desenvolvedores e/ou usuários envolvidos na avaliação. Indica necessidade de treinamento.
 (2) Especialistas são necessários numa comparação entre interfaces, na qual a avaliação tem perfil eminentemente heurística.
 (3) Este consumo de tempo pode ser reduzido com o emprego de equipamentos para automatizar processos na equipe.
 (4) O consumo de tempo pode ser reduzido aumentando o número de observadores, realizando-se sessões simultâneas.
 (5) Necessidade de gravadores de áudio/vídeo, câmaras, dispositivos para *log* e para preenchimento de formulários.
 (6) A utilização de cenários implica na análise de tarefas.
 (7) O *Walkthrough* Cognitivo procura identificar aspectos na interface que venham a representar dificuldades e deficiências cognitivas ao usuário. A partir deste resultado, na etapa de Interpretação de Resultados, são inferidos os problemas de utilizabilidade.

Tabela 7: Aplicabilidade e Adaptabilidade dos Métodos de Avaliação de Interfaces

Sugestão da Adaptação a Problemas Específicos						
Métodos Problemas	Avaliação Heurística	<i>Guidelines</i>	Comparação	<i>Walkthrough</i> Cognitivo	<i>Walkthrough</i> Pluralístico	Teste de Utilizabilidade
Diálogo (*)	2	2	3	1	1	1
Consistência (*)	1	1	1	1	2	2
<i>Look and Feel</i> (*)	1	2	2	1	3	1
Detecção de Problemas de Escopo Geral (*)	2	1	2	2	2	1
Detecção de Problemas de Alta Recorrência (*)	2	1	2	2	2	1
Detecção de Problemas Graves (**)	1	2	3	2	2	1
Quantidade de Problemas Pouco Graves (**)	1	2	1	2	1	3
(*) A taxaço utilizada é: 1 = Bom 2 = Regular 3 = Insuficiente.						
(**) A taxaço utilizada é: 1 = Quantidade Elevada 2 = Quantidade Média 3 = Quantidade Baixa						

Tabela 8: Adaptação a Tipos Específicos de Problemas

5. Conclusão

Apesar de não envolver experimentos, as constatações deste estudo preliminar mostram que os métodos abordados apresentam grandes variações quanto a seus aspectos de facilidade de aplicação, adaptação, efetividade e eficiência. Isto leva a crer que a aplicação destes métodos, as definições necessárias na escolha de determinado método e o planejamento da avaliação são questões de compromisso, verificação de necessidades, objetivos e recursos disponíveis.

Os métodos são propostos para cobrir de maneira eficiente o conjunto de condições que cercam os interessados na avaliação, mas não são imutáveis. Definir uma abordagem própria de avaliação de interfaces pode ser uma solução prática e adequada quando métodos, aplicados isoladamente, não atendem às expectativas de quem está interessado na avaliação. Os métodos podem ser vistos como complementares nestes casos e este estudo tenta fornecer um panorama de alguns métodos utilizados em situações reais para permitir a criação de uma estratégia que se mostre efetiva e eficiente.

6. Referências

- [Dumas 1989] Joseph S. Dumas: “Stimulating Change Through Usability Testing”. In *ACM SIGCHI Bulletin*, July 1989, volume 21, number 1, pp. 37-44.
- [Gomoll 1990] Kathleen Gomoll: “Some Techniques for Observing Users”. In Brenda Laurel, editor, *The Art Of Human-Computer Interface Design*, pp. 85-90, Addison-Wesley Publishing Company, 1993, USA.
- [Jeffries e Desurvire 1992] Robin Jeffries, Heather Desurvire: “Usability Testing vs. Heuristic Evaluation: Was There a Contest?”. In *ACM SIGCHI Bulletin*, October 1992, volume 24, number 4, pp. 39-41.
- [Jeffries et al. 1991] Robin Jeffries, James R. Miller, Cathleen Wharton, Kathy M. Uyeda: “User Interface Evaluation in the Real World: A Comparison of Four Techniques”. In *ACM CHI 1991 Conference Proceedings*, pp. 119-124.
- [Karat et al. 1992] Clare-Marie Karat, Robert Campbell, Tarra Fiegel: “Comparison of Empirical Testing and Walkthrough Methods in User Interface Evaluation”. In *ACM CHI 1992 Conference Proceedings*, pp. 297-404.
- [Nielsen 1992] Jakob Nielsen: “Finding Usability Problems Through Heuristic Evaluation”. In *ACM CHI 1992 Conference Proceedings*, pp. 373-380.
- [Nielsen e Molich 1989] Jakob Nielsen, Rolf Molich: “Teaching User Interface Design Based on Usability Engineering”. In *ACM SIGCHI Bulletin*, July 1989, volume 21, number 1, pp. 45-48.
- [Norman 1986] Donald A. Norman: “Cognitive Engineering”. In Donald A. Norman and S.W. Draper, editors, *User Centered System Design: New Perspectives on Human-Computer Interaction*, pp. 31-62, Lawrence Erlbaum Associate Publishers, Hillsdale, New Jersey, USA.
- [Retting 1994] Marc Retting: “Prototyping for Tiny Fingers”. In *Communications of the ACM*, April 1994, volume 37, number 4, pp. 21-27.
- [Rowley e Rhoades 1992] David E. Rowley, David G. Rhoades: “The Cognitive Jogthrough: A Fast-Paced User Interface Evaluation Procedure”. In *ACM CHI 1992 Conference Proceedings*, pp. 389-396.
- [Smith e Mosier 1986] Sidney L. Smith, Jane N. Mosier: “*Guidelines for Designing User Interface Software*”. Report MTR-10090, The MITRE Corporation, Bedford, Massachusetts, USA, 1986.

- [Treu 1994] Siegfried Treu: *“User Interface Evaluation: A Structured Approach”*. Plenum Press, USA, 1994.
- [Wharton et al. 1992] Cathleen Wharton, Janice Bradford, Robin Jeffries: “Applying Cognitive Walkthrough to More Complex User Interfaces: Experiences, Issues, and Recommendations”. In *ACM CHI 1992 Conference Proceedings*, pp. 381-388.

Agradecimentos

Agradecemos ao professor Dr. Rogério Drummond por ter colaborado na revisão deste relatório, bem como ao CNPq e à FAPESP pelo apoio financeiro oferecido.