

PROPOSTA DE ABORDAGEM PARA SELEÇÃO DE ESTRATÉGIAS AVALIATÓRIAS DE PROCESSOS INTERATIVOS

José Eustáquio R. de Queiroz. Msc

Tecnologista Sênior do Instituto Nacional de Pesquisas Espaciais - INPE

Doutorando do curso de Pós-Graduação em Eng Elétrica

Laboratório de Meteorologia, Recursos Hídricos e Sensoriamento Remoto da Paraíba

Av. Aprígio Veloso, 882 - Bodocongó

Campina Grande - PB Brasil CEP 58.109-970

e-mail: *rangel@lmrs.br*

Tel: (083) 333-2355/Fax: (083) 333-2035

Maria de Fátima Q. V. Turnell. PhD

Professora do Curso de Pós-graduação em Engenharia Elétrica - DEE/UFPB

Coordenação de Pós-graduação em Engenharia Elétrica

Universidade Federal da Paraíba

Av. Aprígio Veloso, 882 - Bodocongó

Campina Grande - PB Brasil CEP 58.109-970

e-mail : *turnellm@dee.ufpb.br*

Tel/Fax: (083) 310-1015 ou.

RESUMO

Muitos esforços ainda serão necessários para justificar os custos dos processos avaliatórios, face aos benefícios que podem advir de sua aplicação. Diante da diversidade de estratégias para avaliação de interfaces torna-se difícil a escolha daquela que poderia levar aos resultados mais efetivos com a melhor relação custo-benefício. Este trabalho apresenta um conjunto de aspectos relevantes à escolha de uma estratégia de avaliação, procura introduzir as intrincadas relações existentes entre eles, e sugere uma abordagem para escolha de uma ou um conjunto de estratégias. Com o propósito de esclarecer a abordagem proposta é apresentado um exemplo de sua aplicação. O trabalho conclui com a proposta de um ensaio avaliatório envolvendo um conjunto de estratégias na avaliação de um mesmo produto, que sirva como referencial de comparação entre elas.

PALAVRAS-CHAVES:

Interfaces homem-máquina, Estratégias de Avaliação, Engenharia de Software

1 Considerações Introdutórias

Desde meados da década passada, inúmeros estudos vêm procurando enfatizar a significância da comunicação entre aplicações e usuários no contexto dos processos interativos usuário-sistema e a importância da *usabilidade* em procedimentos avaliatórios de processos interativos usuário-computador (e.g., [Spen85], [Whit88], [Kara92], [Niel93], [Niel94]). Assim, novas abordagens de desenvolvimento e avaliação de processos interativos, baseadas em modelos mentais do usuário, na estruturação cognitiva de novas informações pelo cérebro humano e em processos de trabalho do usuário vêm sendo postas em prática. Assim como novos procedimentos de avaliação, fundamentados pela engenharia da usabilidade, vêm sendo aplicados junto a comunidades usuárias. Apesar de todos os esforços envidados pelos grupos de pesquisa de processos interativos usuário-computador, conquanto as interfaces de usuário venham adquirindo cada vez mais novas características, há muito poucos modos sistemáticos de torná-las melhores ou mesmo de saber quão eficientes ou deficientes elas são. É quase impossível comparar

projetos - por um lado, porque as medidas de usabilidade podem não estar relacionadas às características de projeto identificáveis; por outro, porque geralmente se deseja comparar dois ou mais projetos para dispensar o esforço de implementá-los todos (ou o embaraço de implementar a opção menos adequada).

Segundo Thimbleby [Thim94], conquanto as interfaces de usuário venham adquirindo cada vez mais novas características, há muito poucos modos sistemáticos de torná-las melhores ou mesmo de saber quão eficientes ou deficientes elas são. É quase impossível comparar projetos - por um lado, porque as medidas de usabilidade podem não estar relacionadas às características de projeto identificáveis; por outro, porque geralmente se deseja comparar dois ou mais projetos para dispensar o esforço de implementá-los todos (ou o embaraço de implementar a opção menos adequada).

Dentre os métodos avaliatórios que têm sido propostos e testados, duas grandes categorias podem ser identificadas: (i) *métodos de avaliação da usabilidade*, desenvolvidos desde meados dos anos 80, frutos dos esforços de inúmeros pesquisadores, dentre os quais Spence [Spen85], Bennett [Benn86], Whiteside [Whit88], Nielsen e Molich [Niel90], Nielsen [Niel93, Niel94], Shneiderman [Shne93] e Karat [Kara92]; e (ii) *métodos de avaliação empíricos e outros*, desenvolvidos e/ou validados por diversos outros pesquisadores, tais como Jeffries e Rosenberg [Jeff87], Somborg [Somb87], Callahan [Call88], Card *et al.* [Card90], Svendsen [Sven91], Morgan *et al.* [Morg91], Gray *et al.* [Gray92], Eberts e Bittianda [Eber93], Benbasat e Todd [Benb93] e Trumbly *et al.* [Trum93].

A motivação-chave deste trabalho reside na constatação de uma carência de estudos comparativos de processos avaliatórios que possibilitem o estabelecimento de critérios de seleção de uma abordagem ou um conjunto de abordagens de avaliação que se adequem a realidades particulares. Um estudo revisivo desta natureza apresenta como contribuição principal a descrição de uma sistemática de seleção de uma ou de um conjunto de estratégias avaliatórias dentre um universo abrangente, segundo um conjunto de critérios pré-definidos.

A seção 2 discute aspectos relativos à seleção de estratégias avaliatórias de processos interativos, sob a forma de problemas explícitos e implícitos, apresentando em seguida a descrição de uma sistemática de seleção e finalizando com um exemplo de aplicação da sistemática. A seção 3 apresenta considerações relativas às limitações da sistemática proposta e à continuidade de seu refinamento.

2 Descrição da Sistemática de Seleção de Estratégias Avaliatórias de Processos Interativos Usuário-Computador

O ponto de partida para a proposição da sistemática ora descrita foi um estudo revisivo abrangente dos diferentes métodos avaliatórios consagrados na literatura. No entanto, a classificação dos métodos estudados, apresentada a seguir, se baseia sobretudo na categorização geral de Treu [Treu94], por ser mais extensiva. Esta seção apresenta, inicialmente, um estudo comparativo das vantagens e desvantagens das diferentes estratégias avaliatórias estudadas, que serve de respaldo para o enquadramento de uma estratégia avaliatória ou uma combinação delas a contextos particulares, à luz da sistemática descrita posteriormente.

O estudo revisivo realizado visou evidenciar o caráter suplementar de cada método estudado com relação ao restante dos métodos, sobretudo no enfoque de aspectos diferentes da engenharia da usabilidade ou outros aspectos do processo interativo usuário-computador, no caso das estratégias empíricas. Além do mais, buscou-se enfatizar vantagens e desvantagens apresentadas por cada uma dessas estratégias sobre as demais. O número de usuários de teste disponíveis para a realização das tarefas inerentes ao contexto avaliatório ou, no caso da avaliação heurística, do número de especialistas disponíveis, bem como o grau de experiência da equipe de avaliação aparecem neste trabalho como variáveis de grande influência na seleção do método ou da combinação de métodos a adotar.

Das duas grandes categorias mencionadas na seção anterior foram estudados:

1 Métodos de Avaliação da Usabilidade

- 1.1 Ensaios de Usabilidade [Niel93, Treu94]
 - I. Observações;
 - II. Uso de Questionários;
 - III. Entrevistas;
 - IV. Verbalização de Procedimentos;
 - V. Interação Construtiva;
 - VI. Ensaio Retrospectivo;
 - VII. Captura Automática diretamente da aplicação;
 - VIII. Discussões em Grupo;
 - IX. Retorno Imediato de Opiniões do Usuário.
- 1.2 Revisões Sistemáticas Cognitivas [Jeff91, Kara92]
- 1.3 Uso de Padrões e Diretrizes de Projeto [Shne93, Quei94]
- 1.4 Avaliação Heurística [Niel93]

2 Avaliação Empírica e Outros Métodos [Spen85, Call88, Whit88, Card90, Gray92, Benb93, Eber93, Treu94]

- I. Representação Comportamental do Usuário e Modelagem Analítica de Processos Interativos;
- II. Comparação de Técnicas e Estilos de Interação com relação a Variáveis Seleccionadas de Desempenho Empregadas na Representação de Amostras de Usuários;
- III. Comparação de Características Alternativas de uma Técnica de Interação Específica;
- IV. Efeito do Uso de Ícones e da Interação Icônica em Processos Interativos Homem-Máquina;
- V. Efeitos de Características Específicas de Interfaces sobre o Desempenho de Usuários;
- VI. Efeitos de Outros Tipos de Dispositivos de Entrada ou Modos de Entrada da Informação (Voz e Toque, Dispositivos Multidimensionais, Digitação por Toque com Caneta Ótica) no Desempenho de Usuários
- VII. Estudo de Relações Gerais entre Usuários e Sistemas Computacionais no tocante à Qualificação e seus Efeitos no Desempenho de Usuários.

A Tabela 1 traça um perfil sumarizado dos métodos de avaliação acima apresentados. Embora construída de um modo bastante sucinto, a fim de facilitar procedimentos de consulta, apresenta aspectos relevantes pertinentes a cada um dos métodos avaliatórios mencionados, tais como estágio do ciclo de vida do produto em que o método se aplica, número típico de usuários de teste, vantagens e desvantagens inerentes. A ausência do item Avaliação Empírica e Outros deveu-se ao fato da diversidade de categorias avaliatórias estudadas e seu extenso leque de propósitos, procedimentos, benefícios e limitações, que dificultou bastante a generalização de aspectos como estágio do ciclo de vida do produto em que se aplica, principais vantagens e desvantagens, e número de usuários de teste necessários.

Embora a descrição de qualquer experimento avaliatório bem sucedido atraia a atenção de quem deseje definir e adotar uma estratégia de avaliação de projetos específicos e, até mesmo, induza o leitor a imaginar que este ou aquele método se adequa aos seus propósitos, não se pode garantir que o método elegido lhe propiciará resultados piores, similares ou melhores do que aqueles obtidos por um autor ou experimentador em outras circunstâncias. Apesar de muito haver sido escrito a respeito, no fundo a impressão que fica após a leitura de um documento típico de apresentação e discussão de estratégias avaliatórias é, por um lado, de dúvida, diante de tantas opções com resultados aparentemente similares, em virtude do modo como as idéias são normalmente abordadas; e, por outro, de impotência, face ao número de questões sem respostas quase sempre imprescindíveis para uma tomada de decisão. Uma reflexão mais aprofundada do assunto conduz à visualização de uma série de problemas que uma tomada de decisão desse nível pode acarretar.

2.1 Problemas Implícitos

Na maioria dos casos, o(s) autor(es) de uma estratégia metodológica ou posteriores experimentadores omitem, em suas descrições, informações que lhes parecem óbvias, por conta da familiaridade com o contexto metodológico. No entanto, tais informações se afiguram geralmente imprescindíveis para quem pretende adotá-la pela primeira vez. Talvez seja este um dos principais fatores que tornam diversas organizações de desenvolvimento de *software*, assim como inúmeros profissionais da área, relutantes em divisar a importância e os benefícios conseqüentes da formação

de equipes encarregadas da estruturação de ambientes dedicados à avaliação de projetos de interfaces e da condução de testes visando a otimização de seus produtos.

Outro problema que merece destaque neste contexto é a falta de ponderação dos investimentos e dos retornos quando da seleção de estratégias metodológicas face à especificidade do contexto da avaliação. Os processos avaliatórios visam, em geral, a otimização da qualidade de um produto e a conseqüente satisfação de seus consumidores. Entretanto, esta é apenas uma das facetas da questão. A perseguição desenfreada da qualidade de um produto pode implicar, sem dúvida, a otimização e o refinamento de suas características técnicas, porém não implica, conseqüentemente, vantagens do ponto de vista econômico[Fava96]. Assim, é imperativa a necessidade de ponderação de aspectos relativos ao grau de otimização da qualidade de um produto face a outras variáveis, e.g., ordem de grandeza do par tempo-custos econômicos associados, retorno do investimento, quotação de mercado, resultados a curto prazo e tendências do mercado competitivo.

Em nível empresarial, [Bera96] também aponta como problemas: (i) a subestimação dos esforços de teste, a partir da redução da etapa de avaliação do produto; (ii) o impedimento de um entrosamento mais efetivo entre as equipes de avaliação e desenvolvimento ao longo de todo o ciclo de vida do produto; e (iii) o incentivo econômico preferencial à equipe de desenvolvimento, a partir da desproporcionalidade do número de contratações de projetistas face ao de avaliadores.

2.2 Problemas Explícitos

Problemas explícitos, no âmbito desta seção, referem-se a decisões da equipe de avaliação, face à diversidade de opções disponíveis. A tomada de decisões, no tocante à opção metodológica mais adequada às circunstâncias consideradas, depende de diversos aspectos da estratégia a ser adotada, de modo que se possa conciliar questões puramente relacionadas com a otimização do produto e aspectos envolvendo relações de custos/benefícios. Dentre os diversos aspectos que norteiam tal tomada de decisões, podem ser encarados como fundamentais: (i) Recursos Necessários; (ii) Tempo Necessário; (iii) Número Mínimo de Avaliadores Envolvidos; (iv) Tipo de Registro; (v) Custo Estimado; (vi) Classe de Problemas Encontrados; e (vii) Percentual de Problemas Encontrados. Infelizmente, de todas as fontes bibliográficas consultadas, apenas a publicação de Jeffries *et al.* [Jeff91] apresenta os resultados de um estudo comparativo de quatro estratégias avaliatórias - ensaio de usabilidade, uso de diretrizes de projeto, avaliação heurística e revisão cognitiva sistemática (*cognitive walkthrough*) - sob um ponto de vista que se aproxima dos critérios acima comentados. Nenhuma das outras fontes revisadas dá indicativos do tempo necessário para conclusão de todo o processo avaliatório proposto/empregado ou dos custos envolvidos, nem tampouco da classe de problemas identificados ou da eficácia percentual do processo avaliatório. Mesmo que o fizesse, ficaria bastante difícil ponderar as indicações de diferentes autores, tanto em virtude da subjetividade dos aspectos quanto pelo fato de que os processos avaliatórios foram aplicados a situações distintas.

2.3 Descrição da Sistemática Proposta

O que se constata, a partir de leituras de diversas referências bibliográficas é que, embora sejam detalhados o ponto de partida, os objetivos e os resultados atingidos com a estratégia avaliatória adotada, o(s) autor(es) não inclui(m) indicações de que a eficácia de tal procedimento metodológico possa ser similar em outras circunstâncias, nem dos esforços dispendidos (em todos os níveis do processo) para alcançá-la. Daí, mesmo que se conclua que o contexto do ensaio se assemelha àquele que se tem em mãos, as lacunas deixadas pela ausência de tais informações lançam dúvidas até sobre a viabilidade da aplicação daquela estratégia a uma realidade diferente. O que se pode prever, quando da adoção de tais procedimentos na íntegra, não contribui muito para a tomada de decisões: é pouco provável que a aplicação de um procedimento metodológico recomendado em realidades distintas possa gerar resultados melhores do que aqueles do experimento original ou é possível que venha a produzir resultados similares. Não se pode prever claramente, sobretudo quando ainda se é principiante no domínio da avaliação de interfaces, se há muita ou pouca chance de que os resultados estejam abaixo do aceitável, i.e., o experimento avaliatório seja um fiasco.

Diante de tais lacunas, surgiu a idéia de conceber, através de uma representação tabular, uma apresentação sumarizada das diferentes estratégias pesquisadas e revisadas neste documento em função dos aspectos discutidos na seção anterior. Não se pode deixar de reforçar a atenção para o fato de que a subjetividade de todos os itens com relação ao empreendimento avaliatório dificulta bastante o estabelecimento de faixas precisas de indicadores semânticos e numéricos que individualizem cada estratégia avaliatória apresentada, sobretudo porque, como já foi frisado anteriormente, a maioria dos ensaios revisados referem-se a aplicativos com graus diferentes de complexidade, o que torna impraticável o estudo comparativo da eficácia, assim como dos recursos necessários em todos os âmbitos (pessoal, material, temporal e econômico). Além disso, cada aspecto está vinculado aos demais através de relações muitas vezes bastante complexas, o que torna o ato de mapeá-los a partir da análises de experimentos distintos uma tarefa bastante difícil e com resultados pouco precisos.

O Quadro 1 comenta alguns dos aspectos considerados, a saber: Recursos Necessários, Número Mínimo de Avaliadores, Classe de Problemas Encontrados e Custo Estimado. Comentários sobre demais aspectos foram omitidos por serem mais evidentes quando da consulta à Tabela 2. A Tabela 2 consta de 8 (oito) colunas, a saber: Método, Recursos Necessários, Tempo Necessário, Número Mínimo de Avaliadores, Tipo de Registro, Custo Estimado, Classe de Problemas Encontrados e Percentual de Problemas Encontrados, nesta ordem. Sua interpretação no estágio atual deve levar em conta de que se trata de um esboço preliminar indicativo, a ser refinado e depurado posteriormente. Todas as faixas são inferidas a partir da projeção dos resultados do trabalho de Jeffries *et al.* [Jeff91]. Eis porque, no estágio atual, esta tabela ainda deva ser consultada com cautela e ponderação. Diante de tais observações, pode-se questionar o motivo de sua apresentação neste documento. Sua contribuição consiste no registro da carência de artifícios de auxílio à tomada de decisões no que concerne à escolha de uma estratégia avaliatória (ou a combinação de um conjunto delas). Este trabalho objetiva documentar o esboço de uma sistemática que se pretende estender e refinar para torná-la um instrumento que possa servir, principalmente aos avaliadores principiantes, como eixo condutor para a estruturação de um ensaio avaliatório no tocante ao aspecto custos-benefícios associados.

QUADRO 1 - Comentários sobre alguns dos aspectos contidos na Tabela 2

ASPECTO	COMENTÁRIOS
Recursos Necessários	● Básico e indispensável (e.g., <i>lápiz e papel</i>) ○ Necessário, mas substituível (e.g., <i>gravador de áudio</i> em entrevistas: substituível por recursos de registro em vídeo) ○ Importante, mas prescindível (e.g., o <i>uso de vídeo</i> incrementa o processo de coleta e revisão de informações, embora seja dispensável de acordo com as circunstâncias)
Número Mínimo de Avaliadores	Idéia de mínimo necessário de competência profissional para a realização da avaliação. A dimensão e complexidade do projeto avaliado influirão na decisão de uma equipe de avaliação mais ou menos numerosa.
Classe de Problemas Encontrados	CLASSIFICAÇÃO PROPOSTA: • Superficiais ⇔ não comprometem a execução das ações tentadas, embora causem desconforto ao usuário, exigindo deste apenas um processo de adaptação, o que não implica que não devam ser solucionados. • Intermediários ⇔ além de causarem desconforto no usuário, o forçam a mudar o curso de suas ações para atingir o propósito almejado, e.g., o acesso "falho" via menu, em um sistema com dois níveis diferentes de interação - menus e manipulação direta - destinados a categorias distintas de usuários [Quei94]. • Graves ⇔ comprometem seriamente a execução de uma linha de ação, e.g., o travamento de uma função que exija o reinício do processo (<i>falha parcial</i>) ou de todo o sistema (<i>falha total</i>). CLASSIFICAÇÃO COMPLEMENTAR [Jeff91]: • De consistência ⇔ relativos a conflitos entre partes do sistema analisado, tanto em nível estrutural e estético quanto semântico e operacional, e.g., mensagens de erro referentes a mesma função do aplicativo ou funções afins apresentadas em diferentes regiões da tela em cores distintas (estrutura e estética), semanticamente divergentes (semântica). • Recorrentes ⇔ interferem na interação a cada vez que se repetem determinadas condições operacionais, e.g., a "quebra" do retorno de informações ao usuário cada vez que este interage com um aplicativo. • Gerais ⇔ ocorrência afeta várias partes do sistema, e.g., falhas que induzem panes parciais ou totais no sistema.
Custo Estimado	Os valores registrados nessa coluna referem-se ao uso de recursos mais sofisticados, e.g., ao considerar as <i>discussões em grupo</i> , deve-se atentar para o fato de que US 2.000,00 seria um custo exorbitante caso se considerasse o recurso e o tipo de registro mais simples - <i>lápiz e papel</i> e <i>relatório escrito</i> , respectivamente. o que já não ocorre quando se pensa no uso do vídeo como recurso e na apresentação final dos resultados sob a forma de relatório ilustrado registrado em fita de vídeo a partir de recursos de editoração eletrônica. Estes valores não incluem os custos com avaliadores.

2.4 Exemplo de Aplicação da Sistemática Proposta

Cada um dos métodos apresentados neste documento apresenta diferentes extensões e limitações, o que só vem reforçar a necessidade, em diversas situações do cotidiano, de combinar alcances e limitações de duas ou mais estratégias avaliatórias numa abordagem adaptativa ou multi-métodos [Quei94]. Ao invés de tentar descobrir um método de avaliação ótimo para uma determinada circunstância, o avaliador poderá empregar seu conhecimento e capacidade de apreciação de metodologias para discernir, dentre a gama disponível de métodos, aqueles que forneçam informações complementares, ampliando a extensão das vantagens e minorando os pontos fracos da abordagem integrada resultante [Lea88].. O exemplo apresentado a seguir ilustra o processo de associação de abordagens metodológicas para a concepção de uma estratégia adaptativa respaldada pelas informações contidas nas Tabelas 1 e 2. O Quadro 2 sumariza os principais aspectos do caso em estudo.

QUADRO 2 - Características Gerais do Exemplo Considerado

ASPECTO	CONSIDERAÇÕES
Motivação	Reclamações verbais de usuários após sessões de uso da aplicação
Alvo da Avaliação	Interface Usuário-Aplicação
Aplicação	Sistema Interativo de Tratamento de Imagens - SITIM-340 v. 2.3
Fase do Ciclo de Vida do Projeto	Atualização e Reprojet
Natureza da Interface	Orientação ao caractere
Mecanismos de Interação	Menus (<i>usuários inexperientes</i>) e comandos (<i>usuários experientes</i>)
Classes de problemas previstas	Gerais, recorrentes e de consistência superficiais e intermediários
Recursos disponíveis	Lápis e papel
Acesso ao universo usuário	Difícil, devido à esparsa distribuição do número relativamente reduzido de organizações operando com o produto em máquinas, na maioria dos casos, não conectadas em rede
Tempo disponível	02 (dois) meses
Nº de Avaliadores disponíveis	02 (dois)
Tipo de Registro circunstancial	Lápis e Papel
Recursos financeiros disponíveis	Nenhum

O fluxograma da Fig. 1 descreve de modo sucinto o processo de seleção de estratégias julgadas as mais convenientes para as circunstâncias acima delineadas - uso de questionários e avaliação heurística - respaldado pelas informações sintetizadas nas Tabelas 1 e 2.

3 Considerações Finais

Este trabalho procurou dar uma visão geral das intrincadas relações existentes entre o grande número de aspectos envolvidos no ato de avaliar, da diversidade de possibilidades de estratégias avaliatórias e, sobretudo, da dimensão do que ainda se pode realizar neste domínio. A sistemática ora apresentada oferece, além do auxílio a processos de tomada de decisão quanto a abordagens metodológicas avaliatórias, auxílio ao processo de concepção de estratégias adaptadas ou derivadas de uma estratégia específica consagrada na literatura ou ainda da hibridação de um conjunto de estratégias. Além disso, também poderá servir de respaldo a outros estudos afins. A sistemática proposta visou evidenciar o caráter suplementar de cada método estudado com relação aos demais, sobretudo no enfoque de aspectos diferentes da engenharia da usabilidade ou outros aspectos do processo interativo usuário-computador, no caso das estratégias empíricas. Além do mais, buscou-se enfatizar vantagens e desvantagens apresentadas por cada uma dessas estratégias sobre as demais. O número de usuários de teste disponíveis para a realização das tarefas inerentes ao contexto avaliatório ou, no caso da avaliação heurística, o número de especialistas disponíveis, bem como o grau de experiência da equipe de avaliação aparecem neste trabalho como variáveis de grande influência na seleção do método ou da combinação de métodos a adotar.

Ficou explícita nas últimas seções a limitação do emprego da Tabela 2: a inferência das faixas apresentadas a partir dos esforços comparativos de Jeffries *et al.* [Jeff91], concentrados apenas sobre 4 (quatro) estratégias metodológicas, como documentado ao longo da seção 4. Deste modo, este trabalho terá continuidade

com a realização de um ensaio avaliatório envolvendo todas as abordagens metodológicas aqui discutidas para a identificação de problemas relativos a um mesmo aplicativo. Atualmente, encontra-se em fase de estruturação um ensaio nesses moldes.

Referências Bibliográficas

- [Benb93] BENBASAT, I. & TODD, P., An experimental investigation of interface design alternatives: Icons vs. text and direct manipulation vs. menus, **International Journal of Man-Machine Studies**, 38, pp. 369-402.
- [Bera96] BERARD, EDWARD V., Bringing Testing Into the Fold, **IEEE Software**, 13(3), May 1996, pp. 91-92 e 95.
- [Call88] CALLAHAN, J.; HOPKINS, D.; WEISER, M. & SHNEIDERMAN, B., An empirical comparison of pie vs. linear menus. In: ACM CHI'88 Conference, 1988. **Proceedings**, Palo Alto, 1988, pp. 95-100.
- [Card90] CARD, S.; MORAN, T. P. & NEWELL, A., The Keystroke-Level Model for User Performance Time with Interactive Systems. In: J. Preece, & L. Keller, **Human-Computer Interaction**, Prentice Hall International (UK) Ltd. - Hertfordshire, 1990, pp. 327-353.
- [Eber93] EBERTS, R. E. & BITTIANDA, K. P., Preferred mental models for direct manipulation and command-based interfaces, **International Journal of Man-Machine Studies**, 38, pp. 769-785.
- [Fava96] FAVARO, JOHN, When the Pursuit of Quality Destroys Value, **IEEE Software**, 13(3), May 1996, pp. 93-95.
- [Gray92] GRAY, W. D.; JOHN, B. E. & ATWOOD, M. E., The precis of Project Ernestine or an overview of a validation of GOMS, 1992. **Proceedings**, Monterey, 1992, pp. 307-312.
- [Jeff87] JEFFRIES, R. & ROSENBERG, J. K., Comparing a Form-based and a Language-based User Interface Evaluation for Instructing a Mail Program. In: ACM CHI + GI'87 Conference, 1987. **Proceedings**, Toronto, 1987, pp. 261-266.
- [Jeff91] JEFFRIES, R.; MILLER, J. R.; WHARTON, C. & UYEDA, K. M., User Interface Evaluation in the Real World: A Comparison of Four Techniques. In: ACM CHI'91 Conference, 1991. **Proceedings**, New Orleans, Louisiana, April 27 - May 2, 1991, pp. 119-124.
- [Kara92] KARAT, CLARE-MARIE; CAMPBELL, ROBERT & FIEGEL, TARRA, Comparison of Empirical Testing and Walkthrough Methods in User Interface Evaluation. In: ACM CHI'92 Conference, 1992. **Proceedings**, Monterey, 1992, pp. 397-404.
- [Kunt96] KUNTZMANN-COMBELLES, ANNIE, Can We Learn Lessons Better?, **IEEE Software**, 13(3), May 1996, p. 21.
- [Lea88] LEA, MARTIN, Evaluating User Interface Designs. In: Tony Rubin (Ed.), **User Interface Design for Computer Systems**, Ellis Horwood Limited, Chichester - England, 1984, pp. 134-167.
- [Morg91] MORGAN, K.; MORRIS, R. L. & GIBBS, S., When does a mouse become a rat? or ... comparing the performance and preferences in direct manipulation and command line environments, **Computing Journal**, 34, pp. 265-271.
- [Niel93] NIELSEN, JAKOB, **Usability Engineering**, Academic Press, Cambridge, MA, 1993.
- [Niel94] NIELSEN, J. & LEVY, J., Measuring Usability, **Communications of the ACM**, 37(4), April 1994, pp. 66-75.
- [Nuse96] NUSENOFF, RON, Software Success: Microsoft's Method, **IEEE Software**, 13(3), May 1996, pp. 107-109.
- [Quei94] QUEIROZ, JOSÉ EUSTÁQUIO R. DE, **Validação de uma Metodologia de Avaliação de Projetos de Interfaces Usuário-Computador**, (Dissertação de Mestrado em Engenharia Elétrica) - DEE/CCT/UFPB, Campina Grande, Julho, 1994.
- [Shne93] SHNEIDERMAN, BEN, **Sparks in Innovation in Human-Computer Interaction**, Ablex Publishing, 1993.
- [Spen85] SPENCE, R. H., **Computer Usability Testing and Evaluation**, Prentice-Hall, Englewood Cliffs, N.J., 1985.
- [Treu94] SIEGFRIED, TREU, **User Interface Evaluation**, Plenum Press, New York, NY, 1994.
- [Trum93] TRUMBLY, J. E.; ARNETT, K. P. & MARTIN, M. P., Performance effect of matching computer interface characteristics and user skill level, **International Journal of Man-Machine Studies**, 38, pp. 713-724.
- [Thim94] THIMBLEBY, HAROLD, Formulating Usability, **SIGCHI Bulletin**, 26(2), April 1994, pp. 59-64.
- [Whit88] WHITESIDE, J.; BENNETT, J. & HOLTZBLATT, K., Usability Engineering: Our Experience and Evolution. In: M. Helander (Ed.), **Handbook of Human-Computer Interaction**, Elsevier Science (North Holland), Amsterdam, 1988, pp. 791-817.

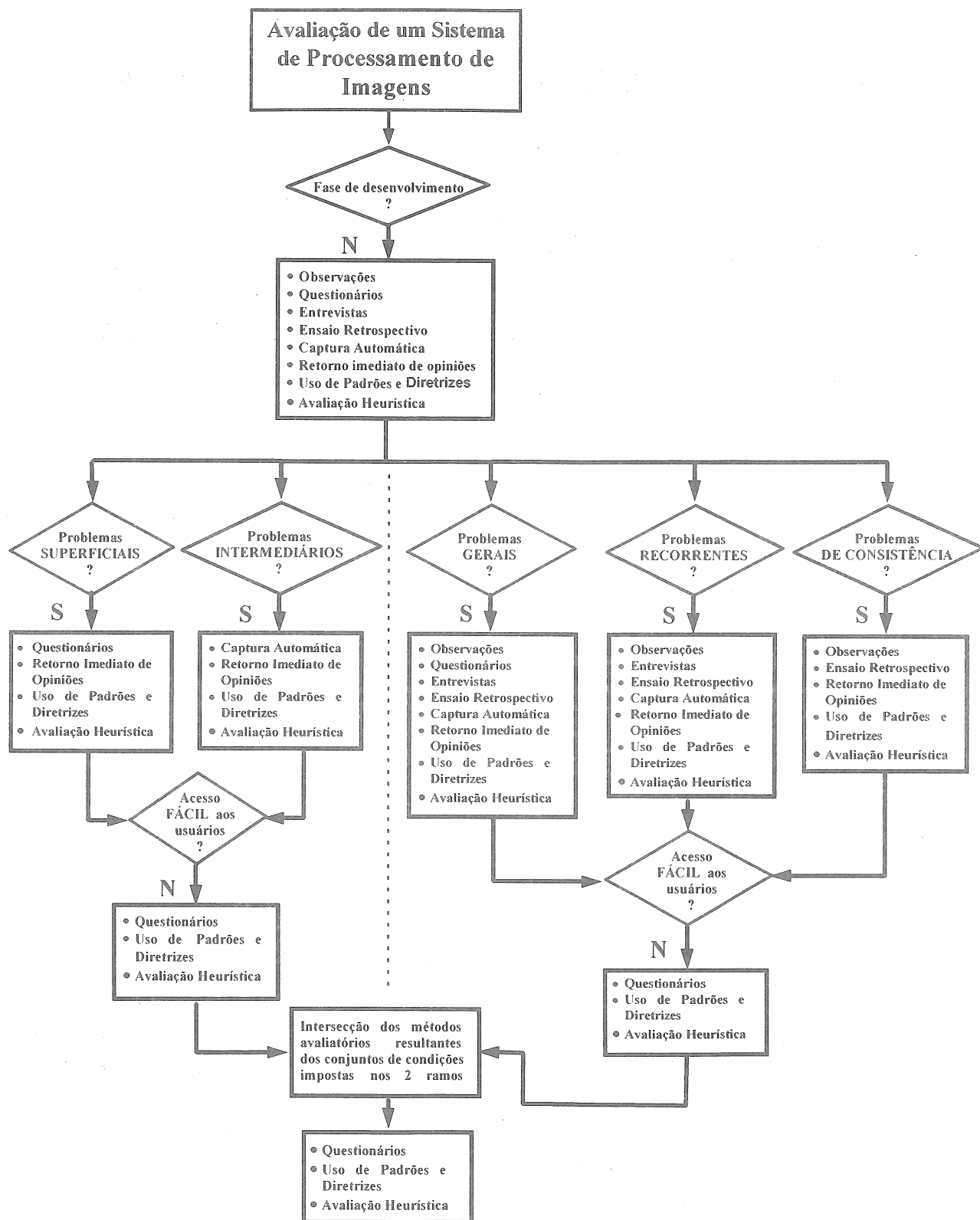


Fig. 1 - Fluxograma do Exemplo de Aplicação da Sistemática Proposta

TABELA 1 - Perfil Sumarizado das Abordagens Metodológicas Revisadas

	Método	Estágio do Ciclo de Vida	Nº de Usuários Necessários	Principais Vantagens	Principais Desvantagens
E n s a i o s d e U s a b i l i d a d e	Observações (Observation)	- Análise de tarefas - Atualização/Reprojeto	03 ou mais	- Verificação da execução de tarefas reais - Facilidade de repetição - Custo relativamente baixo	- Despêndio de tempo - Falta de controle da situação pelo avaliador - Influência da natureza intrusiva do ensaio no desempenho do usuário
	Questionários (Questionnaires)	- Análise de tarefas - Atualização/Reprojeto	Pelo menos 30	- Verificação de preferências subjetivas do usuário - Facilidade de repetição - Custo relativamente baixo - Elevado potencial de público alvo	- Duração total do processo - Baixa taxa de retorno - Dificuldade de elaboração - Dificuldade de análise dos resultados
	Entrevistas (Interviews)	Análise de tarefas	Pelo menos 05	- Verificação aprofundada de atitudes e experiência do usuário - Flexibilidade - Deteção de problemas particulares no usuário	- Despêndio de tempo - Dificuldade de análise e comparação dos resultados - Custo relativamente baixo - Contingência dos resultados à pericia do entrevistador
	Verbalização de Procedimentos (Thinking Aloud)	- Projeto iterativo - Avaliação formativa	03 a 05	- Baixo custo - Direcionamento para concepções errôneas do sistema ou da aplicação pelo usuário - Diagnóstico do aprendizado dos processos interativos pelo usuário	- Fornecimento de resultados pouco representativos para a maioria dos processos de mensuração de desempenho - Caráter intrusivo e não "natural" - Dificuldade de verbalização dos usuários experientes
	Interação Construtiva (Constructive Interaction)	- Projeto iterativo - Avaliação formativa	06 a 10	- Verbalização mais "natural" de procedimentos por crianças/adultos limitados - Aquisição de um maior número de comentários do que o ensaio de verbalização de procedimentos	- Dobro de usuários de teste, com relação às estratégias tradicionais de verbalização de procedimentos - Discrepâncias comportamentais dos usuários de teste, acrescidas de possíveis diferenças de grau de aprendizagem e experiência computacional
	Discussões em Grupo (Focus Groups)	- Análise de tarefas - Projeto cooperativo	06 a 09 por grupo Mais de 01 grupo	- Verificação de reações espontâneas - Prática de dinâmica de grupo para a deteção de problemas - Universalidade dos comentários	- Dificuldade de análise das diversas variáveis envolvidas - Contingência dos resultados à pericia do moderador
	Ensaio Retrospectivo (Retrospective Testing)	- Análise de tarefas - Atualização/Reprojeto	03 ou mais	- Atenuação do caráter invasivo da observação direta - Dispensa de observadores revisito mais extensa e detalhada do comportamento dos usuários de teste - Avanço e retrocesso, além de "congelamento" de cenas previamente registradas em vídeo durante a execução dos testes de usabilidade intencionados	- Pressuposição de um ensaio de usabilidade inicial - Despêndio de tempo - Custo relativamente alto - Inviabilidade em casos de usuários de teste que desempenham atividades críticas (impossibilidade de liberação pela empresa durante a ocorrência dos ensaios)
	Captura Automática direta da aplicação (Automatic Logging from the application)	- Testes finais - Atualização e Reprojeto	Pelo menos 20	- Continuidade do processo - Deteção de recursos da aplicação bastante usados ou muito pouco usados - Custo relativamente baixo	- Violação da privacidade do usuário - Volume de dados gerados - Monitoração exclusiva de ações diretas - Contingência à confiabilidade do hardware instrumentado e do software empregado
	Retorno Imediato de Opiniões do Usuário (User Feedback)	Atualização e Reprojeto	Pelo menos 100	- Monitoração de mudanças nas expectativas e opiniões do usuário - Potencial de coleta de informações de interesse	- Despêndio de tempo na organização e acompanhamento dos resultados - Volume de respostas e/ou informações geradas - Possível contingência a opiniões de técnicos ou especialistas
	Revisões sistemáticas cognitivas (Cognitive Walkthroughs)	- Início de projeto - Ciclo interno de projeto iterativo	Pelo menos 05	- Auxílio à definição de metas e expectativas do usuário - Uso por engenheiros de software - Baixo custo	- Necessidade de metodologia de definição de tarefas - Identificação de muito poucos problemas gerais e recorrentes - Identificação de muito poucos problemas mais sérios - Caráter tedioso e detalhista - Contingência dos resultados à pericia da equipe condutora do teste e ao entrosamento equipe avaliadora/grupo de usuários de teste
Padrões e Diretrizes de Projeto (Standards and Guidelines)		- Início de projeto - Ciclo interno de projeto iterativo - Atualização e Reprojeto	Nenhum	- Identificação de problemas gerais e recorrentes - Uso por engenheiros de software - Baixo custo	- Alto grau de subjetividade das recomendações - Despêndio de tempo para a leitura de coleções volumosas de recomendações - Contingência à experiência do avaliador e à seleção do conjunto de recomendações
	Avaliação Heurística (Heuristic Evaluation)	- Início de projeto - Ciclo interno de projeto iterativo - Atualização e Reprojeto	Nenhum	- Deteção de problemas de usabilidade individuais - Identificação da maioria dos problemas - Identificação de problemas mais sérios - Baixo custo - Informalidade relativa	- Identificação de muito poucos dos problemas mais sérios - Alto custo - Necessidade de vários avaliadores - Não envolvimento de usuários "reais" - Escassez de avaliadores experientes em interação usuário-computador - Identificação de problemas menos sérios, a ser descartados pelos projetistas

TABELA 2 - Critérios de Seleção de Abordagens Metodológicas

Método	Recursos Necessários	Tempo Necessário	Número Mínimo de Avaliadores	Tipo de Registro	Custo Estimado (US)	Classe de Problemas Encontrados	Percentual de Problemas por Avaliador
Observações (Observation)	● Lápis e papel ○ Gravador de áudio ○ Câmera de vídeo	3 semanas	02	● Relatório escrito ● Arquivo em disquete ● Fita de vídeo	2.000,00	Gerais e recorrentes graves	20
Questionários (Questionnaires)	● Lápis e papel	8 semanas	02	● Relatório escrito ● Arquivo em disquete	2.000,00	Gerais e de consistência superficiais	20
Entrevistas (Interviews)	● Lápis e papel ○ Gravador de áudio ○ Câmera de vídeo	3 semanas	02	● Relatório escrito ● Arquivo em disquete ● Fita de vídeo	2.000,00	Gerais e recorrentes graves	20
Verbalização de Procedimentos (Thinking Aloud)	● Lápis e papel ○ Gravador de áudio ○ Câmera de vídeo	2 semanas	02	● Relatório escrito ● Arquivo em disquete ● Fita de vídeo	2.000,00	Gerais e recorrentes graves, intermediários e superficiais	20
Interação Construtiva (Constructive Interaction)	● Lápis e papel ○ Gravador de áudio ○ Câmera de vídeo	2 semanas	02	● Relatório escrito ● Arquivo em disquete ● Fita de vídeo	3.000,00	Gerais e recorrentes intermediários e superficiais	20
Discussões em Grupo (Group Focus)	● Lápis e papel ○ Gravador de áudio ○ Câmera de vídeo	3 semanas	03	● Relatório escrito ● Arquivo em disquete ● Fita de vídeo	2.000,00	Gerais e recorrentes graves, intermediários e superficiais	20
Ensaio Retrospectivo (Retrospective Testing)	● Lápis e papel ○ Câmera de vídeo	1 semana	02	● Relatório escrito ● Arquivo em disquete ● Fita de vídeo	1.000,00	Gerais, recorrentes e de consistência graves intermediários e superficiais (remanescentes)	20
Captura Automática direta da aplicação (Automatic Logging from the application)	● Software de instrumentação do sistema	4 semanas	02	● Relatório escrito ● Arquivo em disquete	1.000,00	Gerais e recorrentes intermediários e graves	20
Retorno Imediato da Opinião do usuário (User Feedback)	● Facilidades de correio eletrônico ○ Software de coleta automática de opiniões	8 semanas	02	● Relatório escrito ● Arquivo em disquete	500,00	Gerais, recorrentes e de consistência superficiais e intermediários	20
Revisões Cognitivas Sistemáticas (Cognitive Walkthroughs)	● Lápis e papel ○ Gravador de áudio ○ Câmera de vídeo	2 semanas	03	● Relatório escrito ● Arquivo em disquete ● Fita de vídeo	3.000,00	Gerais e recorrentes intermediários e superficiais	20
Padrões e Diretrizes de Projeto (Standards and Guidelines)	● Coleções de Diretrizes ● Manual de Regras	Toda a fase de desenvolvimento ¹ 4 semanas ²	02	● Relatório escrito ● Arquivo em disquete ● Fita de vídeo	1.000,00	Gerais, recorrentes e de consistência graves, intermediários e superficiais	20
Avaliação Heurística (Heuristic Evaluation)	● Heurística ● Lista de Inspeção (Checklist)	2 semanas	02	● Relatório escrito ● Arquivo em disquete ● Fita de vídeo	3.000,00	Gerais, recorrentes e de consistência graves, intermediários e superficiais	40

¹ Uso tradicional como guia para desenvolvimento de projetos de interfaces

² Uso para avaliação nos moldes de uma avaliação empírica