

Avaliação em IHC

Departamento de Informática, PUC-Rio

O que é avaliar?

- determinar o valor de
- apreciar
- julgar
- ponderar, examinar, considerar
- calcular, estimar

adaptado de: Dicionário Aurélio - Século XXI em CD-ROM



Avaliação de IHC

- O que é?
 - Examinar aspectos da experiência do usuário, mantendo sempre o foco no usuário. O sistema só interessa na medida em que o seu comportamento afeta a experiência do usuário.
 - Logo, **avaliação de IHC \neq avaliação de software**
- Duas perguntas:
 - Um sistema de busca na Internet é reconhecido como aquele que recupera o maior número de documentos pertinentes para a busca do usuário e o menor número de documentos não-pertinentes.
O que isto informa sobre a experiência que ele oferece aos usuários?
 - Um certo aparelho de telefonia celular é reconhecido como o que permite o maior conforto e a maior rapidez na digitação de mensagens de texto.
O que isto informa sobre a experiência que ele oferece aos usuários?



O que avaliar?

- Aspectos **cognitivos** e **funcionais** relativos à realização de tarefas apoiadas pelo sistema
 - É fácil de aprender? É rápido? Dá para confiar? Dá para reverter erros cometidos (facilmente)? Dá para lembrar depois de algum tempo sem usar? ...
- Aspectos **sócio-culturais** do uso do artefato no local / contexto previstos
 - A tecnologia se integra fácil no ambiente físico? Causa algum tipo de problema com outras tecnologias ou com pessoas que estejam usando outras tecnologias? E com quem não esteja usando tecnologia nenhuma? Pode-se prever mudanças (a curto ou médio prazo) das práticas do ambiente sócio-cultural previsto? Quais? E aí (o que vai provavelmente acontecer)? Há alguma redistribuição de poder? ...
- Aspectos **afetivos**
 - As pessoas vão gostar? Vão achar bonito, agradável? Vão achar que dá status usar a tecnologia (ou vão achar que estão “pagando mico”)? ...



Quando avaliar ?

- Avaliação **formativa** (ou construtiva)
 - ao longo de todo o processo de design
 - artefatos: cenários, *storyboards*, modelagem da interação e protótipos de sistema
- Avaliação **somativa** (ou conclusiva)
 - nas etapas finais (de cada ciclo) do desenvolvimento
 - artefatos: produto (ou protótipo, intermediário ou final) funcionando



Avaliação – objetivos

- verificar o **entendimento** dos projetistas sobre as necessidades e preferências dos usuários
- investigar como uma interface **afeta** a forma de trabalhar dos usuários
- comparar **alternativas** de projeto de interface
- identificar **problemas** de interação ou de interface
 - potenciais (previstos por especialistas) ou reais
- verificar se foram alcançados **objetivos quantificáveis** em métricas de usabilidade
- verificar **conformidade** com um padrão ou conjunto de heurísticas
- elaborar **material de apoio** (e.g. instruções, sistema de ajuda) e treinamento



Paradigma de Avaliação

- Toda avaliação é guiada explícita ou implicitamente por uma série de crenças e expectativas. Várias vezes elas têm origem em teorias.
- Estas crenças e expectativas, juntamente com as teorias a que podem se referir, constituem o paradigma de avaliação.



Paradigmas de Avaliação

- O ‘rápido e rasteiro’
 - Informal, feito a qualquer hora, importante é a rapidez de *feedback*
- O teste de usabilidade
 - Preparação e gravação de sessões de interação, análise criteriosa, muitas vezes combinado a entrevistas e questionários de satisfação
- Os estudos de campo
 - Feito no ambiente do usuário (observação do usuário *in loco*), detecta hábitos e práticas, conscientes/inconscientes
- Avaliações preditivas
 - Previsões de problemas de usabilidade feitas por especialistas, comumente com base em heurísticas, sempre com base em experiência



Técnicas possíveis

- Observação dos usuários
 - em laboratório
 - em seu ambiente de trabalho
- Captura de opiniões
 - dos usuários
 - de especialistas
- Testes de desempenho
- Modelagem do desempenho de usuários ao realizar tarefas (caso da travessia de golfos, por exemplo, modelada por especialista)



Características dos paradigmas de avaliação

	Teste de Usabilidade	Estudos de Campo	Avaliações Preditivas
Usuários	realizam tarefas	comportamento natural	não participam
Local	controlado	natural	qualquer
Quando	protótipo ou produto	início do design	protótipo
Dado	quantitativo	qualitativo	problemas
Retorno para o <i>design</i>	medidas e erros	descrições	problemas
Tipo	Aplicado, baseado em experimentação	Descritivo: objetivo ou etnográfico	especialista



Relação entre Técnicas e Paradigmas

Método	Teste de Usabilidade	Estudos de Campo	Avaliações Preditivas
Observação dos usuários	vídeos, logs	parte central do estudo	n/a
Pedido de opiniões aos usuários	questionários	entrevistas	n/a
Pedidos de opiniões a especialistas	n/a	n/a	heurísticas práticas e experiência
Testes de desempenho	usuários típicos, ambiente controlado	n/a	n/a
Modelagem de Desempenho	n/a	n/a	prever a eficácia, comparar desempenho



Estudos-piloto

- Servem de teste para o “teste pra valer”.
- Visam avaliar
 - O estudo/experimento é viável?
 - Você consegue conduzir o procedimento inteiro e bem?
 - Os scripts, entrevistas, questionários estão claros e sem “bugs”?
- Quantos estudos-piloto se deve fazer?
 - quantos forem necessários para ajustar a qualidade do experimento
 - pelo menos um é indispensável
- E se não tiver usuários “sobrando”?
 - Procurar pessoas com perfil semelhante. Vale até usar seus próprios conhecidos.



Conhecimentos necessários a um avaliador

- Conhecimento sobre o domínio
 - necessário para determinar o que os usuários querem, do que eles precisam, quais são as tarefas mais freqüentes e quais as mais importantes
- Conhecimento sobre o projeto de interfaces de usuário
 - avaliador deve ser capaz de analisar os aspectos mais importantes de um projeto de interfaces (e.g. navegação, terminologia, estruturas de controle, etc.) e conhecer os princípios e diretrizes disponíveis na literatura
- Experiência de utilização do método de avaliação
 - permite ao avaliador ao representar um cliente, bem como o conhecimento sobre o que procurar e o que relatar como resultado da avaliação



Avaliações Analíticas

- Usuários não estão diretamente envolvidos
- Avaliação por inspeção
 - Especialistas fazendo papel de usuários
 - Exs:
 - Avaliação Heurística
 - Walkthrough (percurso) cognitivo
- Modelos preditivos
 - Analisam organização mental e física para realização de tarefas e geram dados quantitativos sobre o tempo que usuário levará para realizar cada tarefa
 - Exs:
 - GOMS, Fitts' Law



Método de Avaliação Heurística

- Características
 - 3 a 5 especialistas em IHC inspecionam uma solução de interface
 - munidos de um conjunto de heurísticas ou diretrizes de usabilidade
 - verificam se a interface viola uma ou mais heurísticas, onde e por quê
 - vantagens: baixo custo, resultados rápidos
 - classificação: avaliação analítica, informal, formativa
- Passos:
 - preparação
 - sessões de avaliação (1-2 horas cada)
 - consolidação
 - seleção dos problemas a serem corrigidos



10 Princípios de Usabilidade (Nielsen 2001)

1. Visibilidade do estado do sistema
 - Mantenha os usuários informados sobre o que está acontecendo, através de feedback adequado, no tempo certo
2. Correspondência entre o sistema e o mundo real
 - Utilize conceitos, vocabulário e processos familiares aos usuários
3. Controle e liberdade do usuário
 - Alternativas, “saídas de emergência”, undo/redo
4. Consistência e padronização
 - Palavras, situações e ações semelhantes devem significar conceitos ou operações semelhantes; caso haja convenções, elas devem ser obedecidas
5. Prevenção de erro
 - Informar usuários sobre consequência de suas ações; impedir ações que levariam a erro



10 Princípios de Usabilidade (Nielsen 2001)

6. Ajuda aos usuários no reconhecimento, diagnóstico e recuperação de erros
 - Mensagens de erro em linguagem simples, sem códigos, indicando precisamente o problema e sugerindo de forma construtiva um caminho para remediar
7. Reconhecimento em vez de memorização
 - Torne objetos, ações e opções visíveis e compreensíveis
8. Flexibilidade e eficiência de uso
 - Ofereça aceleradores e caminhos alternativos para uma mesma ação; permita que usuários customizem ações frequentes
9. Design estético e minimalista
 - Evite porções de informação irrelevante, que reduzem visibilidade relativas das informações relevantes
10. Ajuda e documentação
 - Fáceis de buscar, focadas no domínio e na tarefa do usuário. Devem listar passos concretos a serem efetuados para atingir seus objetivos



Qualificação dos Problemas – Localização

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



Qualificação dos Problemas – Gravidade

- Gravidade (ou Severidade)
 - **frequência** com que o problema ocorre
 - É um problema comum ou raro?
 - **impacto** do problema
 - Será fácil ou difícil para os usuários superarem o problema?
 - **persistência** do problema
 - É um problema que ocorre apenas uma vez e que os usuários conseguem superar facilmente, ou os usuários sofrerão com o problema repetidas vezes?



Escala de Gravidade

4: Catastrófico

- é **imperativo** consertar este problema antes do lançamento do produto
- muitos usuários não conseguirão atingir seus objetivos (para eles, o produto não funciona)

3: Problema grande

- importante de ser consertado; deve receber **alta prioridade**
- muitos usuários ficarão muito insatisfeitos

2: Problema pequeno

- o conserto deste problema é desejável, mas deve receber **baixa prioridade**
- alguns usuários ficarão insatisfeitos em momentos específicos

1: Problema cosmético

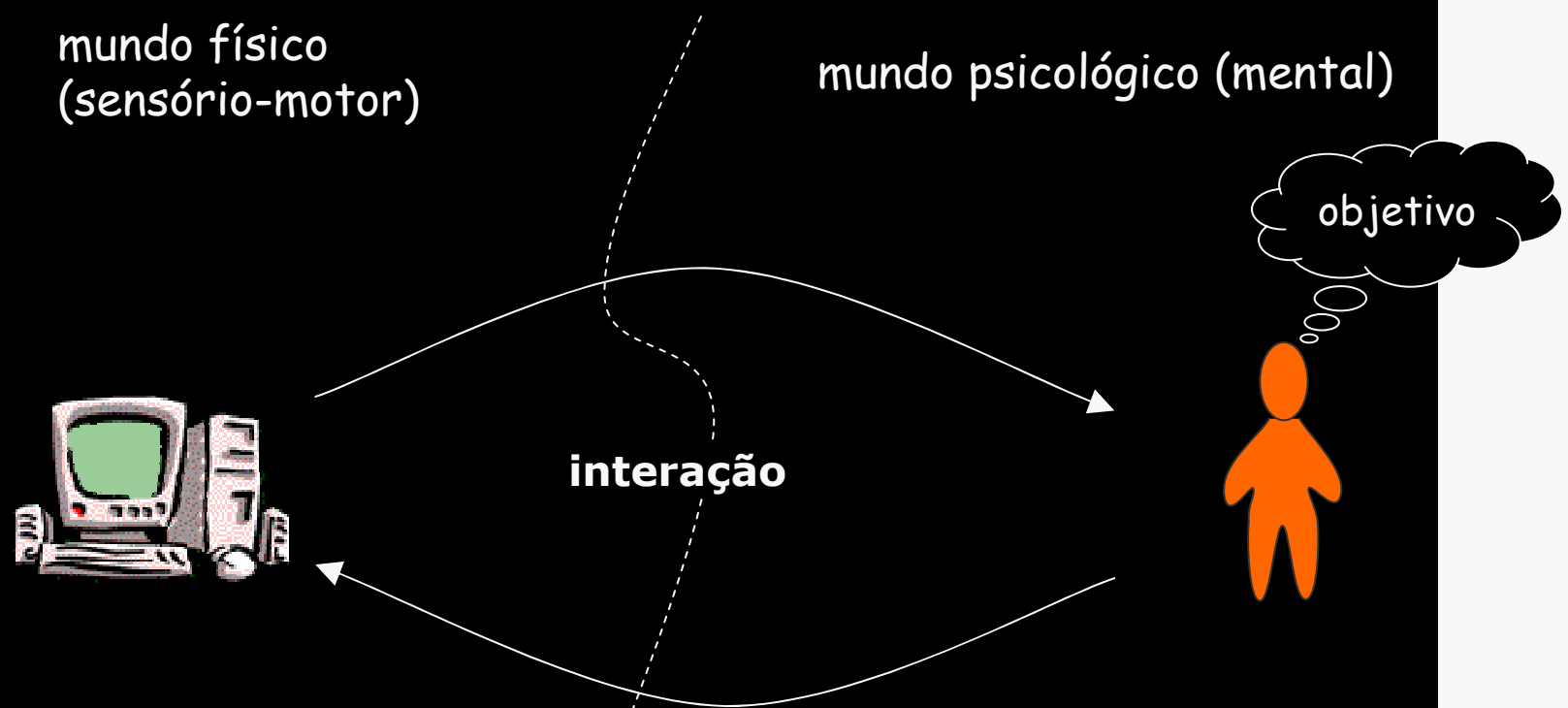


Teoria da ação: 7 estágios de atividades do usuário (Norman, 1986)

- estabelecer o objetivo
- formular a intenção
- especificar a seqüência de ações
- executar as ações
- perceber o estado do sistema
- interpretar o estado do sistema
- avaliar o estado do sistema com relação aos objetivos e intenções

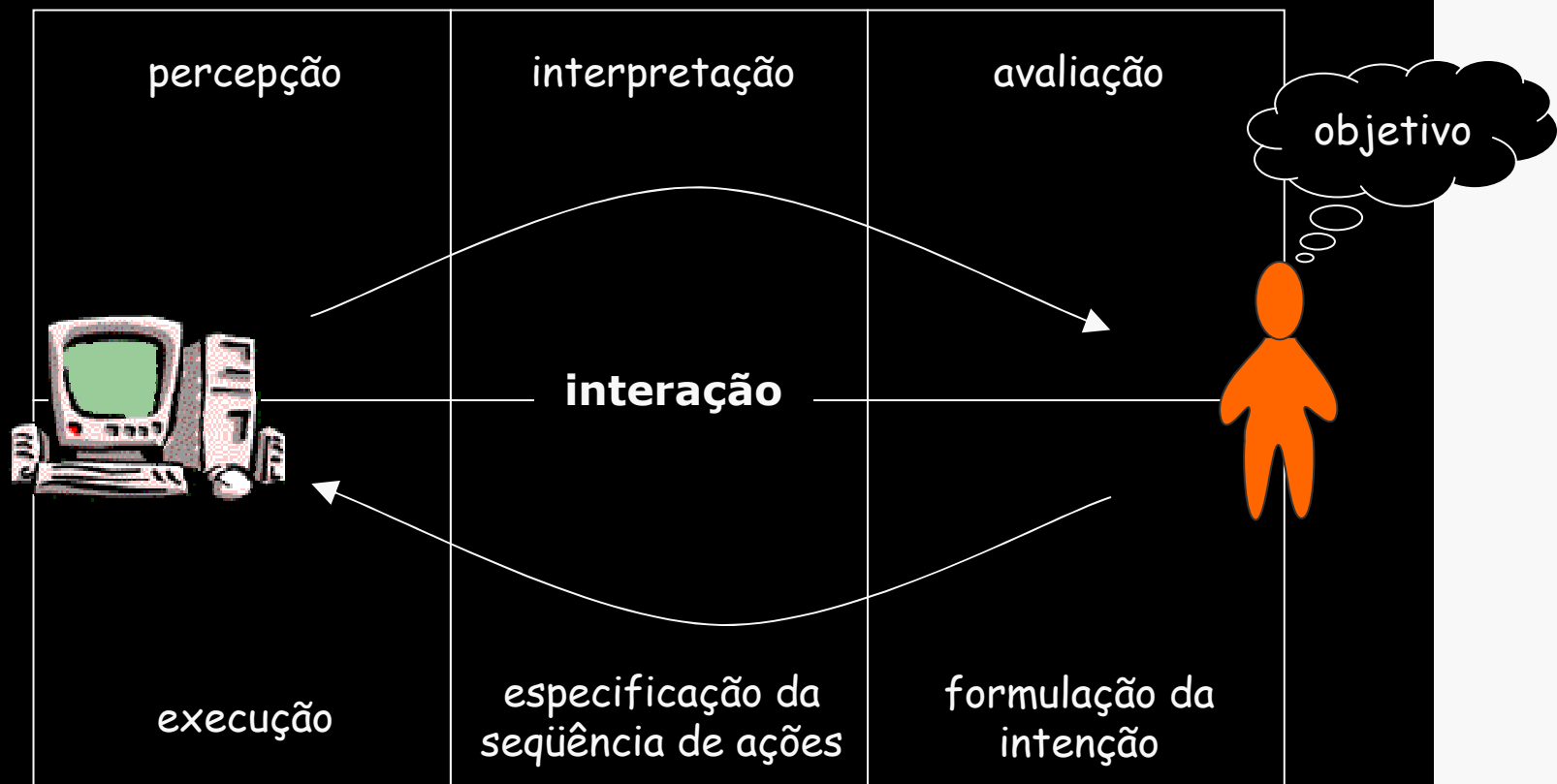


Processo de interação com um artefato



Processo de interação com um artefato

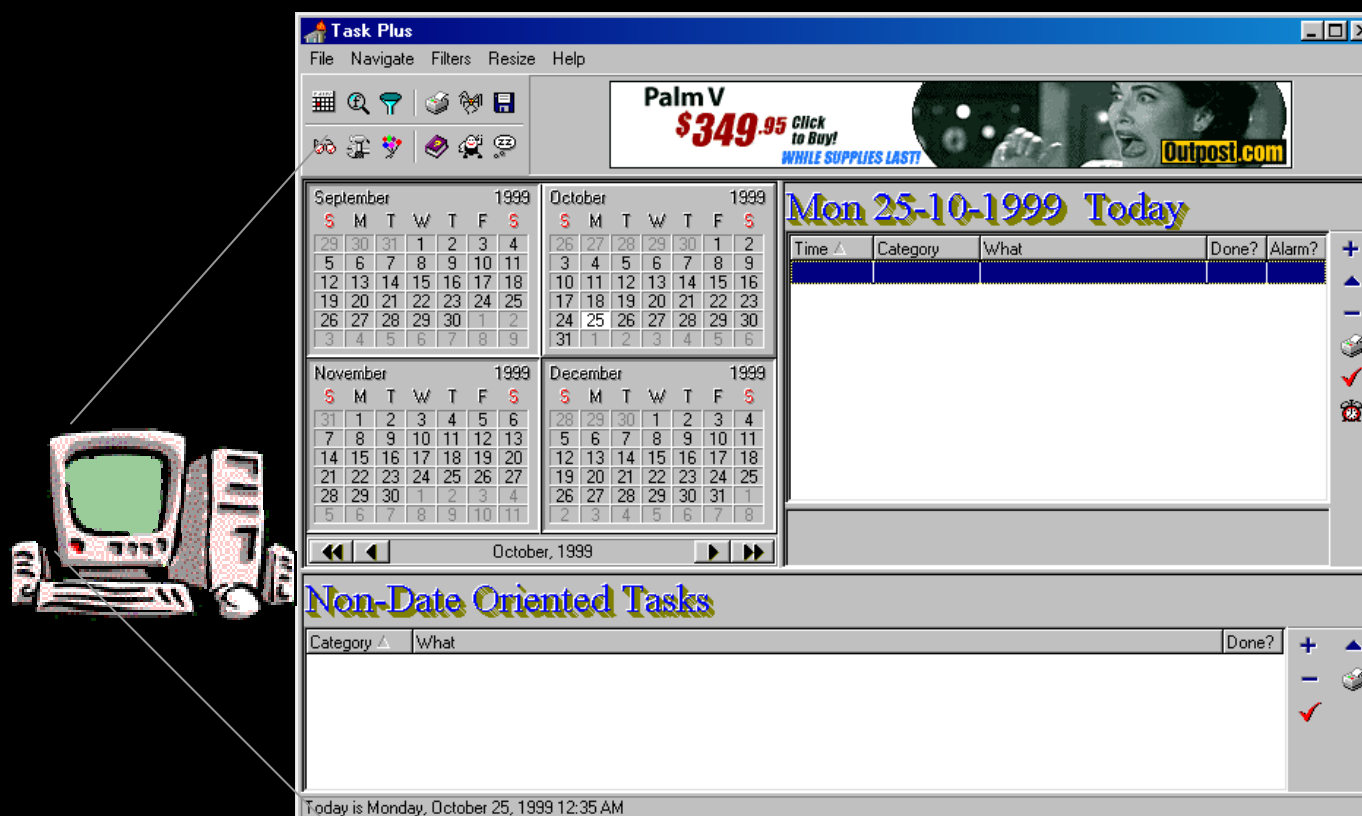
Golfo de Avaliação



Golfo de Execução



Interação com um Sistema de Agenda



Agendar uma reunião.



Exemplo

Golfo de Avaliação

<p>percepção do estado do sistema</p> <p>Apareceu um novo item nos compromissos da 2a. feira.</p>	<p>interpretação</p> <p>Os dados apresentados correspondem à minha reunião.</p>	<p>avaliação</p> <p>Entrei com as informações que eu queria. Completei a tarefa com sucesso.</p>
<p>execução das ações</p> <p>Clico no dia desejado no calendário mensal; Clico em "+ (Add New)"; Digito os dados da reunião no campo "Notes"; Seleciono hora da reunião em "Time"; Clico em "OK"</p>	<p>especificação da sequência de ações</p> <p>Devo selecionar o dia, indicar que quero incluir um compromisso e entrar com os dados que eu quero.</p>	<p>formulação da intenção</p> <p>Quero agendar a reunião de diretoria na 2a. feira da semana que vem.</p>

interação

Agendar uma reunião.



Golfo de Execução



Percurso Cognitivo – Método

- Avaliador
 - projetista, equipe de design ou especialista em IHC
- Preparação para a avaliação:
 - hipóteses sobre a população de usuários (+ contexto)
 - descrição das tarefas representativas a serem executadas
 - lista das ações necessárias para executar cada uma das tarefas
 - proposta de design em papel ou protótipo, ilustrando cada passo e indicando o estado da interface antes/depois de cada um



Percurso Cognitivo – Execução

1. O projetista apresenta uma proposta de design
2. Para cada cenário de tarefas selecionado, o avaliador simula a realização da tarefa passo a passo por um usuário típico. A cada passo, o avaliador...
 - 2.1. realiza uma série de **perguntas** relacionadas à interação de um usuário típico
 - 2.2 constrói histórias plausíveis sobre sucesso ou insucesso da interação com relação a cada pergunta (tentando prever como os usuários agiriam e os problemas que teriam)
 - 2.3 anota pontos-chave, como:
 - o que o usuário precisaria saber antes de realizar a tarefa
 - o que o usuário deveria aprender ao realizar a tarefa



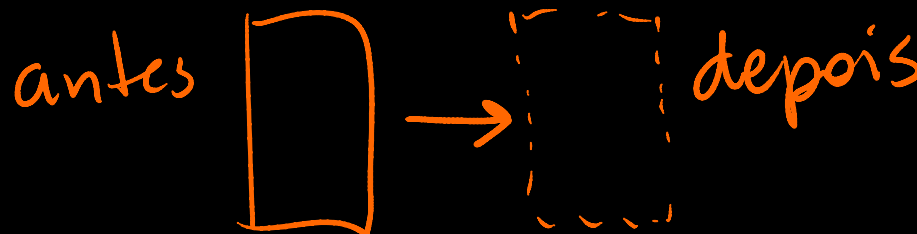
Perguntas sobre cada passo

1) O usuário tentará atingir a meta correta?

- A ação correta estará suficientemente evidente para o usuário? Ele saberá o que fazer para alcançar a meta?
- Dada a decomposição de uma tarefa em subtarefas, o usuário saberá por onde começar? Saberá qual é o próximo passo?
- O que o usuário vai tentar fazer a cada momento?

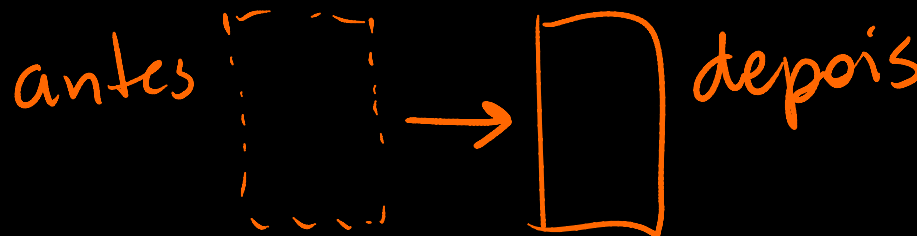
2) O usuário perceberá que a ação correta está disponível na interface?

- Onde está o elemento de interface correspondente ao próximo passo? O usuário pode vê-lo ou sabe onde ele está?
- Que ações a interface torna disponíveis?



Perguntas sobre cada passo (cont.)

- 3) Uma vez encontrado o elemento de interface, o usuário reconhecerá que ele produzirá o efeito desejado?
 - O elemento de interface revela seu propósito e comportamento?
 - O usuário consegue identificar os elementos de interface?
- 4) Após a ação correta ser executada, o usuário perceberá que progrediu em direção à solução da tarefa?
 - Como a interface apresenta o resultado de cada ação?
 - O resultado apresentado tem correspondência com o objetivo do usuário?



Exemplo de formato de relatório para percurso cognitivo

- Breve descrição do sistema
- Perfis de usuário considerados
- Tarefa A – título da tarefa (ex: “Postar” nova mensagem)
 - Sequência correta de ações:
 - 1 – *ação do usuário* (ex: Clicar em “efetuar login”)
 - 2 – *ação do usuário*
 - 3...
 - Avaliação:
 - 1 – *ação do usuário*
 - <tela antes do passo 1> <tela depois do passo 1>
 - Pergunta 1: O usuário tentará atingir a meta correta?
 - » *Sim, porque...*
 - » *Não, porque...*
 - Pergunta 2: O usuário perceberá que a ação correta está disponível na interface?
 - » ...
 - Pergunta 3: Uma vez encontrado o elemento de interface, o usuário reconhecerá que ele produzirá o efeito desejado?
 - » ...
 - Pergunta 4: Após a ação correta ser executada, o usuário perceberá que progrediu em direção à solução da tarefa?
 - » ...
 - 2 – *ação do usuário*
 - <tela antes do passo 2 (= depois do passo 1)>
 - ...



Percurso Cognitivo (resumo das questões)

- 1) Usuários tentarão atingir a meta correta?
 - Dada a decomposição de uma tarefa em subtarefas, o usuário saberá por onde começar? Saberá qual é o próximo passo?
 - O que o usuário vai tentar fazer a cada momento?
- 2) Usuários perceberão que a ação correta está disponível na interface?
 - Onde está o elemento de interface correspondente ao próximo passo?
 - Que ações a interface torna disponíveis?
- 3) Uma vez encontrado o elemento de interface, usuários reconhecerão que ele produzirá o efeito desejado?
 - O elemento de interface revela seu propósito e comportamento?
 - O usuário consegue identificar os elementos de interface?
- 4) Após a ação correta ser executada, o usuário perceberá que progrediu em direção à solução da tarefa?
 - Como a interface apresenta o resultado de cada ação?
 - O resultado apresentado tem correspondência com o objetivo do usuário?



Como avaliar?

- inspeção (por especialistas) – avaliação **analítica**
 - rápida e de baixo custo
 - pode ser realizada ao longo de todo o projeto
- testes com usuários – avaliação **empírica**
 - permite identificar “imprevistos”
 - utilizando
 - maquetes em papel
 - protótipos implementados rapidamente
 - o próprio sistema
 - sistemas concorrentes



Teste de Usabilidade

- Envolve gravação do desempenho de usuários típicos, realizando tarefas típicas.
- Ambiente controlado.
- Usuários são observados e cronometrados.
- Dados são gravados em vídeo e aperto de teclas são logados.
- Os dados são usados para calcular os tempos gastos pelos usuários nas tarefas, bem como para identificar e explicar erros.
- Satisfação do usuário é avaliada por meio de questionários e entrevistas.
- Estudos de campo podem ser usados para prover entendimento contextual.



Experimentos & testes de usabilidade

- Experimentos testam hipóteses para descobrir novos conhecimento através da investigação da relação entre duas coisas ou mais (variáveis).
- Teste de usabilidade é experimentação aplicada.
- Desenvolvedores checam se o sistema é utilizável pela população de usuários prevista, para a realização de suas tarefas.



Testes de usabilidade e Pesquisa

Testes de usabilidade

- Melhorar produtos
- Poucos participantes
- Resultados informam projeto
- Geralmente não replicáveis
- Condições controladas tanto quanto possível
- Procedimento planejado
- Resultados reportados aos desenvolvedores

Experimentos para pesquisa

- Desenvolver conhecimento
- Muitos participantes
- Resultados validados estatisticamente
- Têm que ser replicáveis
- Condições altamente controladas
- Design experimental
- Resultados reportados à comunidade científica



Teste de Usabilidade

- Objetivos & questões focam em quão bem os usuários realizam tarefas com o produto.
- Comparação de produtos ou protótipos são comuns.
- O foco é no tempo para completar tarefas, e também no número e tipos de erros.
- Dados coletados por vídeo e logging de interação.
- Teste é central.
- Questionários & entrevistas provêem dados sobre satisfação dos usuários.



Condições de teste

- Lab de usabilidade ou outro espaço controlado.
- Ênfase em:
 - Selecionar usuários representativos;
 - Desenvolver tarefas representativas.
- Tipicamente seleciona-se 5-10 usuários.
- Tarefas geralmente não demoram mais que 30 minutos.
- Condições de teste devem ser as mesmas para todos os participantes.
- Termos de consentimento informam os procedimentos e lidam com questões éticas.



Alguns tipos de dados coletados

- Tempo pra completar uma tarefa.
- Tempo pra completar uma tarefa após determinado tempo sem usar o produto.
- Número e tipos de erros por tarefa.
- Número de erros por unidade de tempo.
- Número de interações necessárias para help online ou manuais.
- Número de usuários cometendo determinado erro.
- Número de usuários completando a tarefa com sucesso.



Experimentos

- Predizem a relação entre duas ou mais variáveis.
- Variável independente é manipulada pelo pesquisador.
- Variável dependente depende da variável independente.
- Desenho experimental típico tem 1 ou 2 variáveis independentes.



Desenhos experimentais

- Participantes diferentes - grupo único de participantes é alocado aleatoriamente para as condições experimentais.
- Mesmos participantes – todos os participantes aparecem em ambas as condições experimentais.
- Participantes “casados” - participantes são colocados em pares, e.g., baseado em experiência, gênero, etc.



Desenhos Experimentais

Desenho	Vantagens	Desvantagens
Participantes diferentes	Sem efeito de ordem dos experimentos	Muitas pessoas & problema com diferenças individuais
Mesmos participantes	Poucas pessoas, sem diferenças individuais	Precisa algum tipo de contra-balanço por causa do efeito da ordem dos experimentos
Participantes casados	Sem efeito de ordem, com redução das diferenças individuais (com relação aos participantes diferentes)	Não dá pra garantir casamento perfeito em todas as diferenças



Avaliação de Comunicabilidade

- Comunicabilidade
 - Propriedade que bons sistemas interativos devem ter de comunicar com clareza sua lógica de design, ou seja, para que servem, como funcionam, quem são seus usuários visados, etc.
- Objetivo
 - **avaliar a qualidade da (meta)comunicação* entre o designer e os usuários de um sistema computacional**, apontando e explicando as soluções que não deram certo, sugerindo e justificando soluções que podem dar mais certo.

*Metacomunicação: comunicação sobre a comunicação. No caso, a comunicação do designer sobre a comunicação usuário-sistema.



Avaliação de Comunicabilidade

Em linhas gerais, o objetivo é alcançado através da:

1. Análise de como a mensagem está codificada pelos designers;
 - Estudo da documentação *online* e *offline*, e análise dos signos estáticos e dinâmicos.
 - Relação dos problemas identificados por cada avaliador.
 - Elaboração de uma versão da metacomunicação, *i.e.* “reconstrução” da mensagem do designer.
2. Análise das evidências de como ela está “chegando aos usuários” (*i.e.* como os usuários a estão interpretando); e
3. Exploração de situações de metacomunicação não-evidenciadas pelos usuários, mas relacionadas às que foram de fato evidenciadas.



Etapas

- Preliminares
 1. Preparação do teste
 2. Aplicação do teste
 - Embora comuns a testes que envolvem observação de usuários, quando aplicadas na avaliação de comunicabilidade, possuem características peculiares.
- Centrais - Analíticas
 3. Etiquetagem
 4. Interpretação
 5. Elaboração do perfil semiótico





Etiquetagem

- Os avaliadores assistem, individualmente, aos filmes de todos os participantes do teste e, para cada ruptura de comunicação observada, atribuem etiqueta(s) que, para eles, representa(m) a reação do usuário ao que acontece durante a interação.
- As (13) etiquetas são expressões corriqueiras, que “o avaliador coloca na boca do usuário” (como em balões de histórias em quadrinhos).
 - Apesar de corriqueiras, estas expressões têm de ser usadas em sentido técnico.
 - “Cadê?”, “E agora?”, “O que é isto?”, “Oops!”, “Onde estou?”, “Assim não dá”, “Por que não funciona?”, “Ué, o que houve?”, “P’ra mim está bom”, “Desisto”, “Vai de outro jeito”, “Não, obrigado”, “Socorro!”



“Cadê?”

- Usada quando o usuário sabe a operação que deseja executar, mas não a encontra de imediato na interface.
 - Equivale a saber o que dizer, mas não encontrar palavras para dizê-lo.
- Sintomas típicos:
 - Abrir e fechar menus e submenus, listas de *pull-down* ou outras estruturas de signos, passar com o cursor sobre botões, inspecionar diversos elementos de interface, sem ativá-los, à procura de um elemento particular.



(cont.) Etiquetagem: Observações

Início	Fim	Etiqueta	Sintoma	Observação
00:00:07	00:00:27	Cadê?	... Ele lê as opções de configuração da primeira aba, da segunda, da terceira, e volta para a segunda. Está procurando a opção que determina o fechamento automático do tag.	-
00:00:55	00:01:09	Ué, o que houve?	Deixando para trás a estratégia de customização pelo diálogo <i>Preferences</i> , no menu <i>Modify</i> , ele seleciona <i>Modify Current Tag</i> , mas nada acontece. Ele repete a seleção, e de novo nada acontece.	-
...



Esclarecendo a Diferença...

Teste de Usabilidade Teste de Comunicabilidade

- Tipicamente adota a abordagem cognitiva de IHC.
 - IHC é uma atividade cognitiva.
 - Tipicamente tem por objetivo “medir o desempenho de usuários”.
 - Foca a usabilidade.
 - O resultado tipicamente concentra-se no observado.
- Adota a abordagem semiótica, a da Engenharia Semiótica.
 - IHC é metacomunicação.
 - Tem por objetivo ampliar o conhecimento (de desenvolvedores, avaliadores, pesquisadores) sobre como os usuários interpretam o (design do) artefato.
 - Foca a comunicabilidade.
 - O resultado extrapola o dado observado, incorporando inferências e resultados de exploração do próprio avaliador.

Assemelha-se
assim a uma
inspeção...



E a adaptação disso tudo à interação 3D ?

- As heurísticas seriam outras?
- O que medir nos testes de usabilidade?
- Os labels da etiquetagem da avaliação de comunicabilidade são os mesmos?
- Há mais de um designer? (Designer da aplicação, designer do conteúdo 3D, designer da interação 3D)?
- Qual o papel dos aspectos ergonômicos?
- ????
- ????

