

## Capítulo 11

# Planejamento da Avaliação de IHC

### Objetivos do Capítulo

- Discutir a importância de avaliar a qualidade de uso de um sistema interativo.
- Descrever o planejamento e a execução da avaliação de IHC envolvendo ou não usuários.
- Caracterizar os objetivos de avaliação e contextos de projeto que auxiliam na escolha do método de avaliação a ser utilizado.

---

A avaliação de IHC é uma atividade fundamental em qualquer processo de desenvolvimento que busque produzir um sistema interativo com alta qualidade de uso. Ela orienta o avaliador a fazer um *juízo de valor* sobre a qualidade de uso da solução de IHC e a *identificar problemas* na interação e na interface que prejudiquem a experiência particular do usuário durante o uso do sistema. Assim, é possível corrigir os problemas relacionados com a qualidade de uso antes de inserir o sistema interativo no cotidiano dos usuários, seja um sistema novo ou uma nova versão de algum sistema existente.

Com uma visão ampla do processo de desenvolvimento e dos critérios de qualidade desejáveis, este capítulo começa discutindo a importância de avaliar a qualidade de uso de um sistema interativo. Ele descreve o que pode ser avaliado, quando e onde uma avaliação pode ser realizada e quais tipos de dados são coletados e produzidos. Em seguida, descreve os tipos mais comuns de métodos de avaliação e as atividades básicas de uma avaliação de IHC. O capítulo conclui com uma breve apresentação do *framework* DECIDE, proposto para orientar o planejamento de uma avaliação de IHC (Sharp et al., 2007).

#### Grupo 1

### 11.1 Por que Avaliar?

Conhecer critérios de qualidade e seguir processos de fabricação que buscam criar produtos adequados a esses critérios nem sempre resultam em produtos de qualidade. É possível que algo passe despercebido durante a produção e acabe prejudicando a qualidade do produto final. Pode ser um problema de matéria-prima com defeito ou de má qualidade; pode ser um descuido na manipulação de materiais; pode acontecer um erro humano durante o processo de produção, e assim por diante. Em particular, quando estamos trabalhando com sistemas interativos, os problemas costumam ocorrer na coleta, interpretação, processamento e compartilhamento de dados entre os interessados no sistema (*stakeholders*), e até na fase de implementação (por exemplo, um programador pode cometer o equívoco de codificar uma informação errada sobre o domínio). Então, o que é possível fazer para entregar ao consumidor (*i.e.*, usuário final) um produto de qualidade? Além de continuar seguindo processos de design e desenvolvimento comprometidos com a qualidade do produto final, também é preciso avaliar se o produto resultante desse processo atende

aos critérios de qualidade desejados. A avaliação do produto final possibilita entregar um produto com uma garantia maior de qualidade. Para isso, se algum problema for encontrado durante a avaliação, ele deve ser corrigido antes de o produto chegar ao consumidor.

É difícil garantir a “qualidade total” de um produto, porque seria necessário avaliar o produto final em todas as situações de uso possíveis. Além de ser inviável prever todas essas situações, o custo de tal avaliação seria alto demais, e exigiria muito tempo e esforço para sua realização.

Como qualquer produto, um sistema interativo deve ser avaliado sob a perspectiva de quem o concebe, de quem o constrói e de quem o utiliza. Na perspectiva de quem constrói, o objetivo principal da avaliação é verificar se o sistema funciona de acordo com a especificação de requisitos, ou seja, o foco está em verificar se o sistema recebe os dados de entrada, processa e fornece os dados de saída conforme especificado. Os critérios de qualidade avaliados nessa perspectiva estão relacionados à construção de sistemas interativos, tal como robustez e confiabilidade (Avizienis et al., 2004). Esses critérios de qualidade normalmente abstraem o que existe e ocorre fora do sistema (através da interface com usuário durante a interação), e se concentram no que existe e ocorre *dentro* do sistema. Existem vários métodos de avaliação para verificar a qualidade de um sistema interativo do ponto de vista de quem o constrói (Pressman e Maxim, 2014; Delamaro et al., 2007), como testes de caixa-branca e caixa-preta, por exemplo. Essa avaliação costuma ser realizada por uma série de testes ao longo do processo de desenvolvimento, como, por exemplo, testes de unidade, de integração, de sistema e de operação. Mesmo depois de avaliado e confirmado que um sistema interativo funciona conforme sua especificação, ainda assim ele pode apresentar problemas relacionados ao uso, pois o que ocorre fora da interface durante a interação ainda não foi avaliado do ponto de vista do usuário. Em outras palavras, problemas de IHC ainda podem continuar existindo depois que os problemas na construção (das funcionalidades) do sistema forem identificados e resolvidos.

#### Grupo 2

**Nas perspectivas de quem** concebe e de quem utiliza um sistema interativo, a avaliação tem por objetivo principal verificar se o sistema apoia adequadamente os usuários a atingirem seus objetivos em um contexto de uso. Nessa perspectiva, o que existe *dentro* do sistema só interessa à medida que determina o comportamento aparente deste (*i.e.*, que emerge através da interface) e afeta a experiência vivenciada pelo usuário durante a interação. O foco passa a ser o que existe e ocorre *da interface com usuário para fora*. Os critérios de qualidade avaliados nessa perspectiva são relacionados ao uso, tais como: usabilidade, experiência do usuário, acessibilidade e comunicabilidade, apresentados no Capítulo 3. Neste capítulo e no próximo, nos concentramos em métodos de avaliação da qualidade de uso propostos na área de IHC.

Um sistema interativo é resultado de decisões de design tomadas com base na interpretação do designer sobre a situação atual, sobre sua identificação e interpretação das necessidades e oportunidades de melhoria, no seu conhecimento sobre soluções para problemas semelhantes ou relacionados e na sua criatividade para conceber novas soluções para o problema específico em questão. Desse modo, o processo de interação usuário-sistema e o comportamento do sistema durante esse processo são regidos pela lógica definida pelo designer. Os usuários podem ou não compreender e concordar com a lógica do designer, podem ou não julgar a solução de IHC apropriada e melhor do que as soluções existentes e, quando tiverem escolha, podem ou não incorporá-la no seu dia a dia. Portanto, o designer não pode presumir que, se ele aprovar uma solução de interação e interface, o usuário também irá aprová-la e fazer uso dela no seu cotidiano. É importante lembrar que os usuários são pessoas diferentes dos designers e desenvolvedores. Eles muito provavelmente sentem, interpretam, pensam e vivem em culturas e sociedades diferentes, com costumes e gostos próprios, e possuem objetivos diferentes (*e.g.*, usar vs. conceber vs. desenvolver um produto).

As diferenças entre quem concebe e quem utiliza uma solução IHC não podem ser desprezadas. É importante que a solução de IHC seja avaliada do ponto de vista dos usuários, preferencialmente com a participação deles durante a avaliação. Aquele que avalia a qualidade de uso defende o ponto de vista e os interesses dos usuários, atuando como uma espécie de advogado deles durante o processo de desenvolvimento (Prates e Barbosa, 2003). Sempre que possível, a avaliação de IHC deve ser conduzida

por avaliadores que não participaram da concepção da solução, pois eles possuem melhores condições de analisar a solução sob um ponto de vista mais neutro, para defender os usuários e não o design concebido.

**Grupo 3**

A qualidade de uso de um sistema interativo sempre será avaliada. Isso não pode ser ignorado. O usuário final sempre avalia o sistema durante sua experiência de uso, tecendo uma opinião sobre ele. É fundamental que a equipe de desenvolvimento assuma a responsabilidade de avaliar a qualidade de uso do sistema antes de ele ser entregue ao usuário. É preciso buscar um entendimento de como os usuários podem vir a utilizar o sistema e se vão apreciá-lo, principalmente quando ele for um sistema inovador resultante principalmente da criatividade do designer (Sharp et al., 2019). Identificar e corrigir os problemas relacionados com a qualidade de uso antes de o sistema ser entregue ao usuário demonstra profissionalismo da equipe de desenvolvimento (Prates e Barbosa, 2003).

Dentre as razões para avaliarmos a qualidade de uso de sistemas computacionais interativos, Tognazzini (2000) destaca as seguintes:

- os problemas de IHC podem ser corrigidos antes e não depois de o produto ser lançado;
- a equipe de desenvolvimento pode se concentrar na solução de problemas reais, em vez de gastar tempo debatendo gostos e preferências particulares de cada membro da equipe a respeito do produto;
- os engenheiros sabem construir um sistema interativo, mas não sabem e não estão em uma posição adequada para discutir sobre a qualidade de uso. Quem será o advogado do usuário para defender seus interesses durante o processo de desenvolvimento?;
- o tempo para colocar o produto no mercado diminui, pois os problemas de IHC são corrigidos desde o início do processo de desenvolvimento, assim que aparecem, exigindo menos tempo e esforço para serem corrigidos;
- identificar e corrigir os problemas de IHC permitem entregar um produto mais robusto, ou seja, a próxima versão corretiva não precisa já começar a ser desenvolvida no momento do lançamento do produto no mercado.

**Grupo 4**

Avaliar a qualidade de uso de sistemas interativos não representa apenas um aumento do custo de desenvolvimento, como alguns gerentes de projeto costumam pensar. O custo de avaliar a qualidade de uso não costuma ser alto quando comparado ao orçamento global de um projeto de desenvolvimento, e principalmente quando consideramos os benefícios significativos e importantes para o sistema (Rubin, 1994; Bias e Mayhew, 2005). A curto prazo, avaliar a qualidade de uso e corrigir os problemas identificados contribuem para aumentar a produtividade dos usuários, diminuir o número e a gravidade dos erros cometidos durante o uso, e aumentar a satisfação dos usuários. A médio e longo prazo, identificar e corrigir os problemas de IHC contribuem para diminuir o custo de treinamento e suporte, e para planejar versões futuras do sistema, pois chamam a atenção da equipe para partes do sistema que podem melhorar, além de indicar outras que podem ser mais exploradas. Um gerente de projeto não pode deixar de usufruir dos benefícios da avaliação da qualidade de uso pelo simples desconhecimento do que é, de quanto custa e de quais são seus benefícios. É possível equilibrar o custo da avaliação de IHC com os benefícios obtidos.

Para obter esses benefícios, uma avaliação de IHC não pode ser realizada simplesmente entregando (um protótipo de) o sistema para alguns usuários utilizarem e aguardando o relato espontâneo de problemas. Pelo contrário, avaliar a qualidade de uso requer um planejamento cuidadoso da avaliação para que não sejam desperdiçados tempo e dinheiro.

Ao planejar uma avaliação de IHC, o avaliador deve decidir o que, quando, onde e como avaliar, bem como os dados a serem coletados e produzidos, além do tipo de método utilizado. Essas questões são importantes para orientar a escolha do método de avaliação, sua execução e apresentação dos resultados. Nas próximas subseções, discutimos cada uma delas.

## Grupo 5

**11.2 O que Avaliar?**

A questão fundamental de uma avaliação de IHC é definir quais são os objetivos da avaliação, a quem eles interessam e por quê. Os objetivos de uma avaliação determinam quais *aspectos relacionados ao uso* do sistema devem ser investigados. Esses objetivos são motivados por requisições, reclamações ou comportamentos de qualquer interessado no sistema (*stakeholders*): usuários, designers, cliente (“dono do sistema”), desenvolvedores, departamento de marketing etc. Por exemplo, os usuários podem demonstrar desinteresse em utilizar o sistema ou fazer reclamações a respeito dele; o designer pode desejar comparar alternativas de design; o cliente pode querer verificar se a alta qualidade de uso é um diferencial do seu produto; os desenvolvedores podem querer examinar se a nova tecnologia empregada no desenvolvimento da interface agrada os usuários; o departamento de marketing pode querer lançar um produto que atenda a necessidades dos usuários ainda não exploradas pelos sistemas atuais; e assim por diante. Portanto, o avaliador deve estar atento a situações como essas para definir os objetivos de uma avaliação de IHC de acordo com os interesses dos *stakeholders*. A decisão sobre o que avaliar orienta o avaliador no planejamento, na execução e na apresentação dos resultados da avaliação de IHC.

É possível avaliar diversos aspectos relacionados ao uso de acordo com os interesses dos *stakeholders*. Os principais aspectos avaliados são (Hix e Hartson, 1993; Rubin, 1994; Nielsen e Mack, 1994; Sharp et al., 2019):

- apropriação de tecnologia pelos usuários, incluindo o sistema computacional a ser avaliado mas não se limitando a ele;
- ideias e alternativas de design;
- conformidade com um padrão;
- problemas na interação e na interface.

## Grupo 6

**Avaliar a apropriação de tecnologia** requer a participação dos usuários para permitir uma melhor compreensão sobre: o contexto em que o sistema avaliado se insere, quais os objetivos e as necessidades dos usuários, como os usuários costumam alcançá-los, em que grau as tecnologias disponíveis satisfazem suas necessidades e preferências e como elas afetam sua vida pessoal e profissional.

A avaliação desse aspecto pode ser realizada em diferentes momentos do processo de design. Pode consistir em um estudo exploratório para apoiar a atividade de análise. Nesse caso, costumamos avaliar sistemas interativos existentes ou outras formas de apoio às atividades dos usuários, buscando julgar se o uso da tecnologia atual é produtivo e identificar necessidades e oportunidades de intervenção. Já ao longo do design, permite à equipe de projeto julgar se existe um consenso entre a equipe de design, os usuários e demais *stakeholders* sobre o que foi aprendido durante a elicitação e análise de requisitos e sobre a qualidade da solução sendo projetada. Para isso, é comum utilizarmos como insumo protótipos do sistema sendo projetado, em geral representados através de esboços de tela ou cenários de uso, conforme apresentado no Capítulo 9.

A avaliação da apropriação de tecnologia também permite compreender os efeitos da introdução de um sistema interativo novo ou reprojeto no cotidiano dos usuários. Para isso, devemos investigar como os usuários realizam suas atividades antes e depois da intervenção com o sistema, a fim de julgar se o (novo) sistema lhes oferece apoio computacional adequado, conforme definido por algum fator que possa ser observado ou medido. Podemos avaliar, por exemplo, se os usuários são capazes de atingir seus objetivos em menos tempo, com um número menor de erros ou sem a necessidade de treinamento prévio, se estão satisfeitos, se eles se sentem motivados a explorar e aprender novas funcionalidades do sistema, se o uso do sistema é agradável e outras opiniões sobre aspectos gerais ou específicos do sistema. Esse tipo de investigação também nos permite identificar motivos que levam os usuários a não incorporarem um

sistema interativo (ou parte dele) no seu cotidiano.

**Grupo 7** A avaliação de ideias e alternativas de design busca comparar diferentes alternativas de solução de acordo com critérios relacionados com o uso e com a construção da interface com usuário. Por exemplo, com relação ao uso podemos avaliar a facilidade de aprendizado, o apoio à recuperação de erros ou o tipo de intervenção pretendido na vida dos usuários, enquanto com relação à construção podemos avaliar o custo e o tempo necessários para o desenvolvimento de cada alternativa e a infraestrutura de hardware necessária para executar cada proposta de solução. Os critérios de análise e dimensões de comparação de alternativas de design devem ser definidos de acordo com os resultados da análise da situação atual, isto é, o avaliador deve considerar o contexto de uso, os objetivos, as necessidades e as preferências dos usuários, como eles costumam satisfazê-los e por que o fazem assim. A avaliação (comparativa ou não) de ideias e alternativas de solução costuma ser realizada de forma rápida e informal durante a atividade de design como parte do ciclo iterativo de concepção da solução final. É comum utilizar protótipos de interface em vários níveis de detalhes nesse tipo de avaliação, mas também é possível comparar soluções de design de IHC prontas, ou seja, concretizadas em outros sistemas interativos existentes. Essa avaliação pode ser realizada com ou sem a participação dos usuários.

Avaliar a **conformidade com um padrão** é importante quando a solução de IHC precisa ter características específicas determinadas por padrões estabelecidos. Por exemplo, pode ser necessário que a solução de IHC esteja de acordo com os padrões do W3C para acessibilidade. Assim, podemos assegurar que usuários com certas limitações físicas não encontrarão barreiras intransponíveis para acessar a interface do sistema e interagir com ele. Também podemos avaliar se uma solução de IHC segue os padrões do ambiente computacional em que será inserida, como, por exemplo, padrões estabelecidos pelos ambientes de trabalho GNOME® e KDE®, ou pelos sistemas operacionais Microsoft Windows® e MacOS®. Se esses padrões forem seguidos, os usuários acostumados com esses ambientes tendem a ter menos dificuldades para realizar operações básicas, tal como redimensionar e fechar uma janela, por exemplo.

**Grupo 8** Além disso, podemos verificar se a solução de IHC está em conformidade com padrões utilizados em domínios específicos, como correio e comércio eletrônico. Os usuários de sistemas desses domínios já estão acostumados a certos termos e a certas formas de realizar determinadas operações. Desse modo, se a solução de IHC estiver próxima das convenções adotadas no domínio do sistema, os usuários tendem a ter menos dificuldades para utilizá-lo porque já estão familiarizados com essas convenções. Uma organização também pode definir padrões (no sentido de normas) para seus sistemas, e exigir que as soluções de IHC propostas estejam de acordo com esses padrões. De maneira geral, verificar a conformidade com padrões contribui para a consistência e coerência entre as soluções de IHC que seguem esses padrões. A avaliação desse aspecto não exige a participação dos usuários.

**Problemas na interação e na interface** são os aspectos mais avaliados na área de IHC. Na avaliação desses aspectos, o avaliador pode contar ou não com a participação dos usuários para coletar dados relacionados ao uso de sistemas interativos. Ele analisa os dados coletados com objetivo de identificar problemas na interação e na interface que prejudiquem a qualidade de uso do sistema. Os problemas identificados costumam ser classificados de acordo com sua gravidade (grau de impacto nocivo), com a frequência em que tendem a ocorrer e com os fatores que compõem os critérios de qualidade de uso prejudicados — usabilidade, experiência do usuário, acessibilidade ou comunicabilidade. Por exemplo, um avaliador pode relatar e justificar um problema que prejudica a facilidade de recordação (fator de usabilidade) e outro que dificulta a localização na interface de onde o usuário pode expressar determinada intenção de comunicação durante o uso (etiqueta “Cadê?” no método de avaliação de comunicabilidade, apresentado na Subseção 12.2.2).

Os objetivos de uma avaliação de IHC precisam ser detalhados em perguntas mais específicas para torná-los operacionais (Sharp et al., 2007). A elaboração dessas perguntas deve considerar os usuários-alvo, suas atividades e os artefatos utilizados, que normalmente incluem uma representação ou protótipo da solução de IHC a ser avaliada. A Tabela 11.1 apresenta exemplos de perguntas associadas aos objetivos

de avaliação de IHC descritos anteriormente.

Tabela 11.1: Exemplos de perguntas que uma avaliação de IHC pode responder.

exemplos de perguntas a serem respondidas para cada objetivo	
Grupo 10	<b>objetivo: analisar a apropriação da tecnologia</b>
	De que maneira os usuários utilizam o sistema? Em que difere do planejado?
	Como o sistema interativo afeta o modo de as pessoas se comunicarem e relacionarem?
	Que variação houve no número de erros cometidos pelos usuários ao utilizarem o novo sistema? E no tempo que levam para atingir seus objetivos? E na sua satisfação com o sistema?
	O quanto os usuários consideram o apoio computacional adequado para auxiliá-los na realização de suas atividades?
	O quanto eles são motivados a explorar novas funcionalidades?
	Quais são os pontos fortes e fracos do sistema, na opinião dos usuários?
	Quais objetivos dos usuários podem ser alcançados através do sistema? E quais não podem? Quais necessidades e desejos foram ou não atendidos?
	A tecnologia disponível pode oferecer maneiras mais interessantes ou eficientes de os usuários atingirem seus objetivos?
	O que é possível modificar no sistema interativo para adequá-lo melhor ao ambiente de trabalho?
Grupo 10	Por que os usuários não incorporaram o sistema no seu cotidiano?
	<b>objetivo: comparar ideias e alternativas de design</b>
	Qual das alternativas é a mais eficiente? Mais fácil de aprender?
	Qual delas pode ser construída em menos tempo?
	De qual delas se espera que tenha um impacto negativo menor ao ser adotada?
	Qual delas torna mais evidente os diferenciais da solução projetada?
	Qual delas os usuários preferem? Por quê?
	<b>objetivo: verificar a conformidade com um padrão</b>
	O sistema está de acordo com os padrões de acessibilidade do W3C?
	A interface segue o padrão do sistema operacional? E da empresa?
Grupo 10	Os termos na interface seguem convenções estabelecidas no domínio?
	<b>objetivo: identificar problemas na interação e interface</b>
	Considerando cada perfil de usuário esperado:
	O usuário consegue operar o sistema?
	Ele atinge seu objetivo? Com quanta eficiência? Em quanto tempo? Após cometer quantos erros?
	Que parte da interface e da interação o deixa insatisfeito?
	Que parte da interface o desmotiva a explorar novas funcionalidades?
	Ele entende o que significa e para que serve cada elemento de interface?
	Ele vai entender o que deve fazer em seguida?

Que problemas de IHC dificultam ou impedem o usuário de alcançar seus objetivos?

Onde esses problemas se manifestam? Com que frequência tendem a ocorrer? Qual é a gravidade desses problemas?

Quais barreiras o usuário encontra para atingir seus objetivos?

Ele tem acesso a todas as informações oferecidas pelo sistema?

A avaliação de qualquer aspecto relacionado ao uso de sistemas computacionais interativos, principalmente os problemas na interação e na interface e a apropriação da tecnologia pelos usuários, também fornece insumos para elaborar material de apoio e de treinamento, tais como: tutoriais, instruções de uso e sistema de ajuda.

#### Grupo 11

### 11.3 Quando Avaliar o Uso de um Sistema?

Os métodos de avaliação de IHC podem ser aplicados em diferentes momentos do processo de desenvolvimento, dependendo dos dados disponíveis sobre a solução de IHC sendo concebida. Desde o início da atividade de design, o designer explora ideias alternativas de intervenção na situação atual. Essas ideias são elaboradas e refinadas através de ciclos de (re)design e avaliação, até o designer chegar a uma solução de IHC que possa ser construída. A avaliação de IHC realizada durante a elaboração da solução, ou seja, *antes* de termos uma solução pronta, é chamada de avaliação formativa ou construtiva. A avaliação de IHC realizada depois de uma solução estar pronta é chamada de avaliação somativa ou conclusiva (Hix e Hartson, 1993; Sharp et al., 2019).

A **avaliação formativa** é realizada ao longo de todo o processo de design para compreender e confirmar a compreensão sobre o que os usuários querem e precisam, e para confirmar se e em que grau a solução sendo concebida atende às necessidades dos usuários com a qualidade de uso esperada. Ela envolve principalmente analisar e comparar ideias e alternativas de design durante a elaboração da solução de IHC. A avaliação formativa permite identificar tão cedo quanto possível problemas na interação e na interface que possam prejudicar a qualidade de uso, quando os custos de correção ainda são baixos. Diversos artefatos que representam partes de uma solução de IHC podem servir de insumo para uma avaliação formativa, tais como: cenários de uso, esboços de tela, *storyboards*, modelagem da interação e protótipos do sistema em diferentes níveis de detalhe e fidelidade com a solução final, conforme apresentado no Capítulo 9.

A **avaliação somativa** é realizada ao final de um processo de design, quando existir uma solução (parcial ou completa) de interação e de interface pronta, de acordo com um escopo definido. A solução de IHC final pode ser representada por um protótipo de média ou alta fidelidade, ou até mesmo pelo sistema interativo implementado. A avaliação somativa julga a qualidade de uso de uma solução de IHC buscando evidências que indiquem que as metas de design foram alcançadas, ou seja, que o produto possui os níveis de qualidade de uso desejados.

No planejamento da avaliação, o avaliador deve identificar em que momento no processo de desenvolvimento a avaliação será realizada. Isso lhe permite saber quais representações da solução de IHC estarão disponíveis ou se o próprio sistema pronto e funcionando estará acessível. Esse conhecimento auxilia na escolha do método de avaliação a ser empregado.

#### Grupo 12

### 11.4 Onde Coletar Dados sobre Experiências de Uso?

A interação usuário-sistema afeta e é afetada pelo *contexto de uso*, que abrange o ambiente físico, social e cultural em que ela ocorre. Em particular, o usuário costuma utilizar outros artefatos em conjunto com o sistema e interagir com outras pessoas enquanto o utiliza. Por exemplo, o usuário pode fazer anotações



em papel; ele tem a liberdade de organizar as informações de um modo particular nos diversos artefatos utilizados (no papel, sobre a mesa, em pastas, gavetas, armários etc.); ele pode consultar informações sobre o domínio que estejam fora do sistema; o telefone pode tocar no meio de uma operação ou alguém pode conversar com o usuário, distraindo sua atenção; o usuário pode passar por situações de pressão e cobrança maior no trabalho; a conexão com a Internet pode falhar; e assim por diante. Todos esses elementos e acontecimentos comuns num contexto real podem afetar o uso de um sistema interativo. Conhecer esses fatores é importante para avaliar a adequação do sistema ao ambiente real em que ele será utilizado.

As avaliações de IHC que envolvem a participação dos usuários podem ser realizadas em *contexto real de uso* ou em *laboratório*. A **avaliação em contexto**, que constitui uma forma de estudo de campo (veja Subseção 7.5.6), aumenta as chances de verificarmos a qualidade de uso da solução de IHC perante um conjunto maior e mais diversificado de situações de uso. Apesar de não ser capaz de analisar todas as situações de uso possíveis, esse tipo de avaliação fornece dados de situações típicas de uso que não seriam percebidos em uma avaliação em laboratório. Ela permite entender melhor como os usuários se apropriam da tecnologia no seu cotidiano e quais problemas podem ocorrer em situações reais de uso. Todavia, é difícil controlar a execução de uma avaliação em contexto para assegurar que certos aspectos do sistema sejam analisados.

Grupo 13

A **avaliação em laboratório** oferece um *controle* maior sobre as interferências do ambiente na interação usuário-sistema e facilita o *registro* de dados das experiências de uso com a solução de IHC avaliada. O laboratório é um ambiente preparado para proporcionar experiências de uso sem interrupções ou inconvenientes que podem ocorrer em um contexto real de uso e até mesmo atrapalhar certos aspectos da avaliação do sistema. Uma avaliação em laboratório permite comparar de forma consistente as experiências que diferentes usuários tiveram com o sistema. Sem as interferências do contexto de uso, o usuário possui melhores condições de manter o foco nas tarefas sendo avaliadas.

Dependendo do método utilizado, uma sala de reunião com mesa e cadeiras pode ser um ambiente adequado para realizar uma avaliação. Esse é o caso dos métodos de grupo de foco e prototipação em papel. Outros métodos de avaliação são mais indicados para ambientes especiais que favoreçam a observação, como, por exemplo, o teste de usabilidade e o método de avaliação de comunicabilidade. Nesses casos, o laboratório costuma ser um ambiente projetado e construído para facilitar a observação e o registro de dados sobre experiências de uso. De qualquer forma, apesar de ser um ambiente artificial, o laboratório deve ser confortável para os participantes da avaliação, conforme visto na Seção 7.4.

Um ambiente de observação de uso costuma possuir duas salas: uma sala onde o usuário vai utilizar o sistema (*sala de uso*) e outra onde o avaliador vai observá-lo através de um vidro espelhado (*sala de observação*). Desse modo, quem está dentro da sala de uso não consegue ver o que as pessoas que estão na sala de observação estão fazendo. Pelo menos um avaliador fica na sala de observação, fazendo anotações e acompanhando a interação do participante através de um monitor clone, que reproduz tudo o que ocorre no monitor do participante. Um outro avaliador pode ficar atrás do participante como apoio, mas buscando não expressar opiniões ou fornecer instruções que prejudiquem ou invalidem a avaliação.

Em geral, apresentamos ao participante a sala de observação para que ele conheça o que fica do outro lado do vidro espelhado, entenda melhor o procedimento de avaliação e, assim, fique mais tranquilo. O objetivo do vidro espelhado é diminuir a interferência das ações dos observadores sobre o comportamento do participante. Desse modo, os avaliadores que estiverem na sala de observação durante a sessão podem fazer anotações livremente ou consultar algum material de apoio, sem com isso influenciar os resultados da avaliação. A Figura 11.1 apresenta um exemplo de *layout* de laboratório para observar a experiência de uso de um participante.

Grupo 14

O **registro dos dados observados** pode ser feito de várias formas, orientado direta ou indiretamente pelo método de avaliação. A sala de uso deve ter microfones e câmeras de vídeo para gravar falas, gestos e expressões do participante, e um computador com software instalado para capturar um vídeo



da interação do participante com o sistema e, ocasionalmente, registrar uma lista das teclas digitadas durante a experiência de uso.

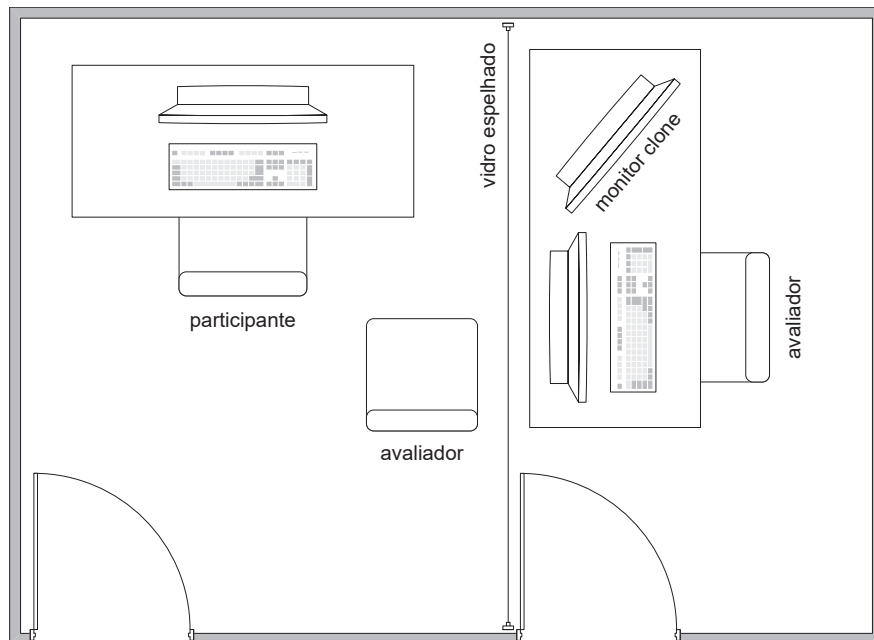


Figura 11.1: Exemplo de laboratório para observar um participante utilizando um sistema computacional interativo.)

Outras configurações de laboratório são possíveis dependendo do sistema avaliado. Por exemplo, se for avaliado um sistema de TV digital interativa, a sala de uso deve ter um sofá e uma TV, em vez de uma mesa com computador pessoal e uma cadeira. Uma outra instalação consiste numa sala para diversos observadores, e uma sala de controle, por exemplo, para avaliações do tipo *Wizard of Oz*, nas quais o participante do teste pensa estar interagindo com o sistema, mas de fato é uma pessoa que está enviando as respostas “do sistema” para o seu terminal.

## 11.5 Que Tipos de Dados Coletar e Produzir?

Dependendo do tipo de avaliação, o avaliador pode coletar dados sobre a situação atual, uso atual da tecnologia, aspectos positivos e negativos identificados durante esse uso, necessidades e oportunidades de intervenção. A abrangência e o foco da coleta de dados devem ser definidos de acordo com os objetivos da avaliação. Por exemplo, se o objetivo for “verificar se determinado sistema (ou protótipo) satisfaz as necessidades dos usuários”, um avaliador poderia coletar dados sobre: o grau de satisfação dos usuários em relação ao sistema; se os usuários sentem necessidade de utilizar outros sistemas ou artefatos para realizarem suas atividades e por que são necessários; quais são os pontos fortes e fracos do sistema avaliado; o que os usuários gostariam que fosse melhorado ou estendido; e assim por diante. Se o objetivo for outro, como, por exemplo, “identificar problemas na interação e interface de determinado sistema (ou protótipo)”, o avaliador poderia coletar dados sobre: quantos usuários conseguiram concluir certas tarefas; quanto tempo foi necessário para concluir cada tarefa; quais erros foram cometidos, em que locais da interface e com que frequência; quantos usuários sentiram necessidade de consultar material de apoio (tutoriais, manuais de uso, ajuda on-line etc.) e em quais momentos da interação o fizeram, quais eram as dúvidas e se foram esclarecidas e assim por diante.

Os dados coletados são interpretados e analisados de acordo com o método de avaliação escolhido, para produzirem resultados que atendam aos objetivos da avaliação, ou seja, que busquem responder as

perguntas específicas elaboradas na definição da avaliação. Cada método descrito no próximo capítulo fornece ao avaliador orientações mais detalhadas sobre quais dados coletar, como analisá-los e quais dados produzir como resultado da avaliação.

Os dados coletados e produzidos em uma avaliação de IHC podem ser classificados de diferentes maneiras. As classificações mais comuns são: nominais, ordinais, de intervalo e de razão (Stevens, 1946; Kiess e Green, 2008; Freedman et al., 2007); dados qualitativos e quantitativos (Hix e Hartson, 1993; Sharp et al., 2019); dados subjetivos e objetivos (Meister e Rabideau, 1965). Como visto ao longo do próximo capítulo, cada método de avaliação de IHC privilegia dados e resultados de diferentes tipos.

Fim - Grupo 15

**Dados nominais ou categóricos** representam conceitos na forma de rótulos ou categorias. Por exemplo, uma pessoa pode ser classificada conforme sua origem étnica em caucasiana, africana, hispânica, asiática etc. Podemos atribuir um número para dar nome à categoria (*e.g.*, 1=caucasiano), mas esse número não transmite qualquer informação quantitativa, apenas serve como identificação dos dados, e poderia ser substituído por letra sem qualquer perda de informação.

Entre dados nominais não existe qualquer relação de ordem. Por exemplo, não se pode dizer que um dado nominal é maior ou menor, melhor ou pior, ou mais alto ou mais baixo do que outro. É possível dizer que os dados nominais são iguais ou diferentes, mas nem sempre se consegue caracterizar o *tipo* de diferença que existe entre eles. Exemplos de dados nominais são: atividades que um usuário realiza no sistema, sistemas semelhantes que o usuário já utilizou, formas de acesso à Internet que o usuário utiliza, idiomas que ele entende e nos quais sabe se expressar e problemas enfrentados durante uma sessão de avaliação.

Dados **ordinais**, como o próprio nome revela, representam conceitos com relações que definem algum tipo de ordem entre eles. Essencialmente, dados ordinais produzem um ranqueamento entre pessoas ou coisas, no qual alguém ou algo possui uma variável em maior quantidade ou intensidade do que outros (Kiess e Green, 2008). Por exemplo, a relação entre o website que um usuário mais utiliza e o segundo site mais utilizado por ele. Embora possamos identificar se um dado ordinal é maior ou menor que outro, não podemos quantificar com precisão a sua diferença. Por exemplo, é possível que, por trás dos dados ordinais sobre os sites A, B e C, em ordem de frequência de uso, esteja o fato de o usuário visitar o site A diversas vezes ao dia, o site B uma vez por dia e o site C a cada 15 ou mais dias. Em outras palavras, ao analisar dois ou mais dados ordinais, é possível dizer *qual* é maior ou melhor, mas não em *quanto*.

Se, além da relação de ordem entre os dados, há uma diferença de igual magnitude entre eles, tratam-se de dados **de intervalo (ou intervalares)**. Eles representam períodos, faixas ou distâncias entre os dados ordinais, mas a origem da escala é arbitrária, ou seja, uma medida de intervalo não possui um valor *zero verdadeiro* para indicar a total ausência da variável medida (Kiess e Green, 2008). Por exemplo, na escala de temperatura em graus Celsius, podemos afirmar que a diferença entre 20°C e 40°C é a mesma que entre 40°C e 60°C. No entanto, não podemos afirmar que a temperatura de 40°C é duas vezes maior do que a temperatura de 20°C. Já um dado **de razão** possui um valor zero verdadeiro, como, por exemplo, o tempo que uma pessoa leva para realizar uma tarefa. Se os participantes P1 e P2 levaram dois e seis minutos para realizar uma tarefa, respectivamente, faz sentido dizer que P2 levou o triplo do tempo de P1 para realizá-las. Também são dados de razão a frequência de acesso à Internet, o número de erros cometidos, o número de contatos que um usuário possui em uma comunidade virtual. Os dados intervalares e de razão podem ser **contínuos** ou **discretos**. Por exemplo, o número de erros cometidos durante a realização de uma tarefa é um dado de razão (possui um zero absoluto — nenhum erro) e discreto (não faz sentido dizer que foi cometido 0,5 erro), já o tempo para a realização da tarefa é um dado de razão e contínuo, pois pode ser fracionado.

Os dados coletados com as escalas frequentemente utilizadas em questionários de IHC, como as escalas de Likert e os diferenciais semânticos (veja Subseção 7.5.2), não são estritamente intervalares. No entanto, costumam ser considerados como dados de intervalo nas estatísticas produzidas, não apenas em IHC, mas também nas ciências comportamentais (Kiess e Green, 2008).

**Grupo 16** **Dados qualitativos representam** conceitos que não são representados numericamente, como as respostas livres coletadas em questionários e entrevistas, representando expectativas, explicações, críticas, sugestões e outros tipos de comentário. Dados nominais também podem ser considerados qualitativos.

Já os dados **quantitativos** representam numericamente uma quantidade, ou seja, uma grandeza resultante de uma contagem ou medição, tais como: o tempo e número de passos necessários para alcançar determinado objetivo; o número de erros cometidos durante uma sessão de uso; quantas vezes a ajuda on-line e o manual de uso foram consultados; e quantos usuários conseguiram alcançar o objetivo satisfatoriamente (Wixon e Wilson, 1997). Nessa classificação se encaixam os dados intervalares e de razão. Os dados ordinais são por vezes classificados como quantitativos, pelo fato de representarem um ranqueamento, ou qualitativos, pois não faz sentido fazer certos cálculos com o número que indica o ranqueamento, como média, por exemplo.

Os dados quantitativos são utilizados com frequência para verificar hipóteses, possivelmente formuladas a partir de uma teoria ou de uma pesquisa qualitativa prévia. Por exemplo, a hipótese de que uma solução A é melhor do que uma solução B poderia ser verificada, dentre outras maneiras, contabilizando quantos usuários conseguem concluir certas tarefas em um tempo determinado utilizando cada solução. Nesse caso, obteríamos resultados do tipo: na solução A, 61% dos usuários concluíram as tarefas com sucesso, 23% deles concluíram metade das tarefas, e 16% não concluíram sequer a metade das tarefas. Já na solução B, os dados coletados foram 82%, 13% e 5%, respectivamente. Analisados isoladamente, esses números não explicam por que alguns usuários não concluíram as tarefas. Também não sugerem recomendações para tornar a solução A ainda mais adequada aos usuários e às suas tarefas. Apenas fornecem evidências de que a solução A é melhor do que a solução B e, caso os resultados sejam estatisticamente significativos, comprova a hipótese do estudo.

Diferente do foco na *contagem e medição* de quantidades realizadas na análise de dados quantitativos, a análise de dados qualitativos envolve principalmente a *interpretação* de conceitos por eles representados. Por isso, ao escolher trabalhar com dados qualitativos, um avaliador normalmente está interessado em *explorar* e *explicar* o que ocorreu (ou pode ocorrer) durante a interação usuário-sistema e como, em vez de testar uma hipótese.

**Fim Grupo 16** **A terceira classificação** que utilizamos distingue dados objetivos e subjetivos. Dados **objetivos** podem ser medidos por instrumentos ou software, como, por exemplo, os termos que um participante utilizou para realizar uma determinada busca (dado objetivo e qualitativo), as músicas que ele mais ouviu no último mês no seu computador (dado objetivo e qualitativo) e o tempo que ele levou para realizar uma tarefa numa sessão de teste (dado objetivo e quantitativo). Já os dados **subjetivos** precisam ser explicitamente expressos pelos participantes da avaliação, como opiniões e preferências. Tais dados podem ser expressos através de textos em linguagem natural (dado subjetivo e qualitativo) ou, em uma escala numérica (dado subjetivo e quantitativo, geralmente coletado através de questionários — veja a Subseção 7.5.2).

Nem todo uso de dados qualitativos consiste na realização de uma pesquisa qualitativa. A **pesquisa qualitativa** consiste em um conjunto de práticas interpretativas e materiais que tornam o mundo visível e transformam-no em uma série de representações, incluindo anotações em campo, entrevistas, conversas, fotografias, gravações e anotações pessoais. A pesquisa qualitativa envolve uma abordagem interpretativa e naturalista do mundo, através da coleção e análise de uma variedade de materiais empíricos que descrevem momentos e significados rotineiros e problemáticos nas vidas dos indivíduos. Esses materiais incluem: estudo de caso, experiência pessoal, introspecção, história de vida, entrevistas, artefatos, textos e produções culturais, textos observacionais, históricos, interacionais e visuais. Os pesquisadores estudam as coisas em seu ambiente natural, buscando interpretar os fenômenos em termos dos significados que as pessoas lhes dão. Eles utilizam uma ampla gama de práticas interpretativas inter-relacionadas, sempre buscando obter um melhor entendimento do assunto em questão (Denzin e Lincoln, 2008, p. 4).

Grupo 17

## 11.6 Qual Tipo de Método de Avaliação Escolher?

Existem vários métodos para avaliar a qualidade de uso propostos na literatura. Cada método atende melhor a certos objetivos de avaliação, orienta explícita ou implicitamente quando e onde os dados devem ser coletados, como eles devem ser analisados, e quais critérios de qualidade de uso (usabilidade, experiência do usuário, acessibilidade ou comunicabilidade) sua análise privilegia. Os métodos de avaliação de IHC podem ser classificados em: métodos de investigação, de observação de uso e de inspeção.

Os métodos de **investigação** (*inquiry*) envolvem o uso de questionários, a realização de entrevistas, grupos de foco e estudos de campo, entre outros. Esses métodos permitem ao avaliador ter acesso, interpretar e analisar concepções, opiniões, expectativas e comportamentos do usuário relacionados com sistemas interativos. Em particular, permitem investigar alternativas de design, problemas que os usuários costumam enfrentar, como eles se apropriaram da tecnologia existente e quais são suas expectativas para futuras interações com tecnologias atuais e novas. São frequentemente utilizados em etapas iniciais do processo de design, para ratificar ou retificar o entendimento da situação atual e identificar necessidades e oportunidades de intervenção (*i.e.*, em análise), bem como explorar formas alternativas de intervenção (*i.e.*, em avaliação formativa). Também são utilizados para avaliar a introdução da nova tecnologia (*i.e.*, em avaliação somativa). Em geral, os métodos de avaliação através de investigação não exigem que os usuários utilizem um sistema interativo durante a coleta de dados. No entanto, o uso de um sistema ou de materiais de apoio, como imagens, cenários e outros tipos de artefatos, pode contribuir para a investigação, ajudando os usuários a se lembrarem de suas experiências e expectativas e também a manterem o foco nas questões de interesse. Os dados obtidos através de investigação com usuários e demais *stakeholders* podem ser coletados através das técnicas discutidas no Capítulo 7, tais como: entrevista, questionário, grupo de foco, estudo de campo e investigação contextual. Vale observar que a qualidade dos dados coletados através de métodos de investigação, como entrevistas, depende fortemente da habilidade do participante em articular suas ideias e em se lembrar de acontecimentos relevantes. Sempre que possível, recomendamos integrar a investigação à observação, de forma a obter dados mais robustos e um entendimento profundo das questões sob investigação.

Grupo 18

Os métodos de **inspeção** permitem ao avaliador examinar (ou inspecionar) uma solução de IHC para tentar *antever* as possíveis consequências de certas decisões de design sobre as experiências de uso. Em outras palavras, tentar identificar problemas que os usuários podem vir a ter quando interagirem com o sistema. Esses métodos geralmente não envolvem diretamente usuários e, portanto, tratam de experiências de uso *potenciais*, e não reais. Além de permitir comparar designs alternativos e buscar problemas em soluções de IHC, os métodos de inspeção permitem ainda avaliar a conformidade com um padrão ou guia de estilo.

Ao inspecionar uma interface, os avaliadores tentam se colocar no lugar de um usuário com determinado perfil, com um certo conhecimento e experiência em algumas atividades. Mas existe um limite na empatia do avaliador; de fato, ele não é o usuário. Ele pode deixar de encontrar problemas que os usuários teriam, e também pode julgar como problemáticos pontos que não causariam dificuldades aos usuários. Além disso, um avaliador pode se concentrar mais em alguns aspectos de usabilidade do que em outros (Nielsen, 1994c). Sempre que possível, devemos buscar dados de diferentes fontes e métodos de avaliação para fazer uma apreciação mais robusta do sistema em questão. Os métodos de avaliação através de inspeção, também denominados métodos analíticos, podem ser utilizados ao longo de todo o processo de design, à medida que modelos ou protótipos são elaborados. Costumam ser mais rápidos e ter um custo menor do que métodos que envolvem usuários. Alguns métodos de inspeção são descritos na Seção 12.1.

Grupo 19

Os métodos de **observação** fornecem dados sobre situações em que os usuários realizam suas atividades, com ou sem apoio de sistemas interativos. Através do registro dos dados observados, esses métodos permitem identificar problemas reais que os usuários enfrentaram durante sua experiência de uso do sistema sendo avaliado. O avaliador pode observar os usuários em contexto ou em laboratório.

A observação em contexto permite coletar uma gama mais ampla de dados mais ricos sobre a atuação dos usuários em seu ambiente de atividade (veja estudos de campo na Subseção 7.5.6). Já a observação em laboratório costuma ser mais direcionada e simples, pois o avaliador tem controle sobre o ambiente. Alguns métodos de avaliação através de observação em laboratório, também denominados de métodos empíricos, são descritos na Seção 12.2.

Métodos de avaliação por inspeção costumam ser mais rápidos e de custo de execução mais baixo do que os métodos de investigação e de observação, pois eles não gastam tempo com recrutamento e sessões de coleta de opiniões ou de observação de usuários. Por exemplo, Salgado et al. (2006) relatam que uma avaliação por inspeção (avaliação heurística) gastou menos da metade do tempo do que uma avaliação com a participação dos usuários (avaliação de comunicabilidade). Entretanto, os resultados de uma avaliação por inspeção são baseados apenas na experiência do avaliador, com base em hipóteses sobre os usuários. Mesmo que o avaliador se coloque no lugar do usuário, ele *não é* o usuário. Não podemos ignorar os limites dessa empatia do avaliador. Apesar de ser necessário mais tempo para coletar e analisar dados empíricos de experiências de uso, os métodos de avaliação através de investigação e observação costumam fornecer resultados mais interessantes do que as previsões dos avaliadores. Isso se deve ao fato de que os usuários percorrem caminhos não previstos pelo avaliador, de forma criativa e oportunista, proporcionando maior realidade, riqueza e diversidade nas experiências de uso analisadas.

**Grupo 20** Os objetivos da avaliação, detalhados por questões específicas, são os guias principais para o avaliador escolher os métodos de avaliação a serem utilizados. Se o objetivo da avaliação for encontrar problemas de IHC, o avaliador pode julgar mais adequado empregar um método por inspeção para cobrir (quase) toda a interface, e selecionar um pequeno número de partes importantes a serem avaliadas por um método de observação ou de investigação. Geralmente, ele selecionaria as partes cuja inspeção não forneceu resultados suficientemente confiáveis. Já avaliar a forma como os usuários se apropriam de tecnologia requer o emprego de um método de avaliação através de investigação ou de observação, por contar com a participação dos usuários. E, para avaliar a conformidade com um padrão, é mais adequado empregar um método de avaliação por inspeção, pois a participação dos usuários é desnecessária.

Além disso, o avaliador precisa escolher métodos de avaliação adequados ao prazo, orçamento, equipamentos, número de usuários acessíveis, número de avaliadores capacitados e experientes em cada método e demais recursos disponíveis. Ao planejar a avaliação, ele tem de verificar se poderá contar com a participação dos usuários para coletar e registrar dados sobre experiências de uso, ou se deverá inspecionar a interface para tentar prevêê-las. Essas questões práticas são responsáveis por viabilizar a execução da avaliação de IHC planejada. Não adianta o avaliador planejar uma avaliação de IHC sem ter condições de executá-la. Por exemplo, o avaliador deve verificar se é possível ter acesso ao contexto de uso com o tempo e a liberdade necessários, para que possa observá-lo adequadamente e coletar e registrar livremente os dados necessários.

**Grupo 21**

## 11.7 Como Avaliar?

Os métodos de avaliação de IHC possuem as seguintes atividades básicas: preparação, coleta de dados, interpretação, consolidação e relato dos resultados. Caso a avaliação encontre problemas ou oportunidades de melhoria, também é planejado um reprojeto do sistema.

### 11.7.1 Por Onde Começar?

Como primeiro passo para preparar uma avaliação, o avaliador deve **aprender** sobre a situação atual, que inclui o domínio do problema, os papéis e perfis dos diferentes grupos de usuários, seus objetivos e atividades, e o contexto em que o sistema é ou será utilizado. O avaliador também deve conhecer as interfaces dos sistemas complementares ou semelhantes com os quais os usuários estejam acostumados a

utilizar, além de, é claro, a interface do próprio sistema ou protótipo a ser avaliado. Sempre que possível, o avaliador deve buscar saber quais são os comportamentos e as dificuldades típicos dos usuários durante o uso de sistemas interativos semelhantes. O Capítulo 7 apresenta algumas técnicas de análise que podem auxiliar os avaliadores nesse aprendizado.

Além de necessário para planejar a avaliação adequadamente, esse entendimento contribui para a coleta e análise dos dados. No caso de uma avaliação por inspeção, ajuda os avaliadores a se colocarem no lugar dos usuários ao tentarem prever problemas na interface e na interação. No caso de uma avaliação através de investigação ou de observação, ajuda os avaliadores a compreenderem certos dados fornecidos pelos usuários.

## Grupo 22

**11.7.2 Preparação**

Apesar de alguns equivocadamente considerarem-na burocrática, a atividade de preparação é fundamental para a condução adequada de uma avaliação que forneça resultados úteis e confiáveis. Ela não pode ser negligenciada, pois pode acarretar em desperdício de tempo, dinheiro e outros recursos, envolvendo avaliadores, usuários e demais interessados na avaliação.

Os **objetivos** da avaliação são definidos com base em requisições, reclamações ou comportamentos dos *stakeholders* do sistema. Se o avaliador conhece o que será avaliado, ele tem melhores condições de compreender as motivações dos interessados e de ajudá-los a definir adequadamente os objetivos da avaliação. Os objetivos devem ser detalhados através de **questões mais específicas** que a avaliação deverá responder, conforme apresentado na Seção 11.2.

Raramente avaliamos o sistema inteiro. Em vez disso, precisamos definir o **escopo** da avaliação, delimitando quais partes da interface, caminhos de interação, tarefas e perfis de usuário devem fazer parte da avaliação. Essa delimitação é feita de acordo com os objetivos e as questões que a avaliação pretende responder. Na escolha das tarefas a serem investigadas, o avaliador pode considerar as tarefas mais importantes para os usuários, as que apresentam mais problemas durante sua realização usando a tecnologia atual, as que foram mais difíceis de projetar, as que motivaram a produção do sistema ou as que são o diferencial do sistema com relação a sistemas semelhantes ou complementares. Além disso, o avaliador deve considerar o prazo e os recursos disponíveis. Uma sessão de teste com usuários, por exemplo, costuma durar em torno de uma hora.

O avaliador **escolhe um ou mais métodos** de acordo com os objetivos da avaliação, dos recursos disponíveis e do acesso aos usuários e ao contexto de uso, conforme apresentado na Seção 11.6. Caso seja escolhido um método de avaliação que envolva usuários, o avaliador deve também **escolher o perfil e o número de participantes**, com base nos perfis de usuários, nos objetivos e no escopo da avaliação. Por exemplo, se o objetivo for verificar como usuários novatos aprendem a realizar determinadas tarefas utilizando o sistema, o avaliador deve recrutar usuários inexperientes no uso do sistema e na realização das tarefas em questão.

## Grupo 23

**Os avaliadores devem buscar** participantes que representem o público-alvo do sistema avaliado, ou seja, que possuam características semelhantes aos usuários típicos. A definição dos perfis de participantes pode considerar fatores como: idade, sexo, formação acadêmica, grau de conhecimento sobre o domínio, nível de experiência na realização das tarefas e nível de experiência no uso do sistema avaliado e de sistemas semelhantes, por exemplo. Sempre que possível e pertinente, o avaliador deve buscar equilibrar o número de homens e mulheres.

Dumas e Redish (1999) relatam que uma avaliação de IHC em geral envolve de cinco a 12 usuários. Nielsen (2000), por sua vez, afirma que bastam cinco usuários para encontrarmos a maioria dos problemas na interface (85%, segundo o seu experimento), alcançando uma boa relação custo-benefício. Caso seja necessário obter resultados estatisticamente significativos, a amostra de usuários deve ser suficientemente grande e representativa. Entretanto, o tempo e outros recursos necessários para a coleta e análise de dados



de muitos usuários pode inviabilizar uma abordagem estatística. Por isso, uma avaliação de IHC com frequência pretende apenas obter indícios sobre a qualidade de uso do sistema e sobre como aumentá-la. Em outras palavras, mesmo quando os resultados não são estatisticamente significativos, eles podem ser úteis para o reprojeto do sistema avaliado.

**Grupo 24** O planejamento de uma avaliação envolve **questões práticas**, que incluem alocar pessoal, recursos e equipamentos e preparar o material de apoio. Uma avaliação com usuários requer também a preparação do ambiente de teste, a realização de um teste-piloto e o recrutamento dos participantes. Ela envolve ainda **questões éticas**, apresentadas na Seção 7.4.

De acordo com os métodos de avaliação escolhidos, o avaliador deve **alocar pessoal, recursos e equipamentos necessários**. Pode ser preciso alocar outros avaliadores que auxiliem na coleta, análise e divulgação dos resultados. Caso os avaliadores tenham pouca experiência com o método selecionado, eles podem precisar de treinamento. Equipamentos para o registro de dados, como câmeras, máquinas fotográficas e gravadores de áudio são comumente utilizados na coleta de dados, e equipamentos mais sofisticados podem requerer a contratação de profissionais especializados. Diversos softwares podem apoiar as diferentes atividades de avaliação, desde a captura de dados até a análise e divulgação dos resultados.

Antes de começar a coletar dados, o avaliador deve **preparar e imprimir o material de apoio necessário**. No caso de avaliações que envolvam participantes, esse material costuma incluir:

- termo de consentimento livre e esclarecido (TCLE), de acordo com os cuidados éticos necessários (veja Seção 7.4);
- questionário pré-teste (ou roteiro de entrevista estruturada) para coletar informações dos participantes que podem influenciar a interação usuário–sistema, tais como: características pessoais, experiências anteriores com tecnologia e conhecimento sobre o domínio;
- roteiro de entrevista pós-teste para coletar informações sobre a opinião e os sentimentos do participante decorrentes da experiência de uso observada;
- instruções e cenários para orientar os participantes sobre as tarefas a serem realizadas;
- roteiro de acompanhamento da observação, de modo a facilitar a captura de dados e anotações.

**Grupo 25** Para assegurar a validade do estudo, é importante que todos os participantes recebam o mesmo material, as mesmas informações e orientações.

Quando for elaborar os cenários de tarefa, o avaliador deve estar atento ao tempo necessário para os participantes realizarem as tarefas descritas. O tempo estimado para um participante realizar cada tarefa não deveria ultrapassar 20 minutos. Se for necessário mais tempo, o participante deve poder fazer uma pausa rápida para evitar que sua fadiga física ou mental interfira no resultado da avaliação. Podemos sugerir, por exemplo, que o participante desvie o olhar para um ponto distante depois de passar algum tempo olhando somente para a tela, ou faça um alongamento das mãos e dos braços depois de usar o teclado e o mouse durante algum tempo. Cada participante normalmente despende cerca de uma hora numa sessão de avaliação. Mais do que isso seria cansativo para ele (Sharp et al., 2019). Caso o avaliador tenha dificuldade de planejar tarefas curtas, ele pode revisar o escopo da avaliação.

O formulário de acompanhamento das sessões de observação facilita a anotação de dados importantes para análise da interação, tais como: falas espontâneas do usuário durante o uso, se o usuário conseguiu ou não concluir uma tarefa, se foi necessário ajudá-lo em uma tarefa para prosseguir com a tarefa seguinte etc. Além de observar e registrar dados durante uma sessão de interação, o avaliador pode usar entrevistas curtas antes e depois da execução das tarefas para coletar outros dados importantes.

O avaliador deve **preparar todo ambiente, hardware e software** necessário para o uso do sistema avaliado e a captura de dados. Todas as instalações, configurações e demais procedimentos de preparação para a sessão de avaliação devem ser concluídos antes de receber cada participante. Devemos configurar



e testar câmeras de vídeo, gravadores de áudio, monitores-clone para o registro e acompanhamento das ações do participante e softwares que capturem diversos dados como, por exemplo, um filme da interação do usuário com o sistema, as teclas digitadas e os movimentos e as ações com o mouse. O laboratório deve estar limpo e arrumado para receber os participantes. Quando os participantes forem utilizar o sistema, ele deve estar funcionando perfeitamente.

**Grupo 26**

**Concluído todo o planejamento** da avaliação, é muito importante que o avaliador **realize um teste-piloto**. O objetivo desse teste é avaliar o próprio planejamento, e analisar se a avaliação, tal como planejada, produz os dados necessários para responder a questões e objetivos do estudo. O avaliador deve conduzir o teste-piloto como se fosse uma sessão normal de avaliação. Dessa forma, ele tem oportunidade de verificar se a linguagem nas explicações e nos materiais fornecidos é clara e objetiva, e se esses materiais contêm informações adequadas e suficientes para orientar o participante durante a avaliação. O teste-piloto também permite ao avaliador verificar se o sistema a ser avaliado, os equipamentos e os softwares para registrar os dados estão funcionando corretamente, e se o tempo necessário para executar as tarefas solicitadas é adequado. Mas o teste-piloto não se limita à atividade de coleta de dados. O avaliador deve fazer uma breve análise dos dados coletados, para se certificar de que eles permitirão responder as questões de avaliação. A escolha de quais dados devem ser coletados deve ser cuidadosa. Devemos registrar apenas dados que possam contribuir com os objetivos da avaliação. Além do desperdício de recursos, registrar dados desnecessários pode tornar a sessão de uso mais artificial e constrangedora para os participantes. Por exemplo, gravar a imagem dos participantes em vídeo pode intimidá-los desnecessariamente, caso o vídeo não seja analisado.

Se algum problema for detectado no teste-piloto, o planejamento da avaliação e o material de apoio devem ser corrigidos. Caso sejam feitas muitas mudanças ou o avaliador sinta necessidade, outros testes-pilotos podem ser executados, até que tudo esteja pronto para a realização da avaliação definitiva. Os dados do teste-piloto devem ser descartados, pois podem estar contaminados por problemas que ocorreram durante o piloto ou não serem compatíveis ou válidos considerando o planejamento revisado da avaliação.

**Grupo 27**

**Caso haja um número** reduzido de pessoas com o perfil desejado disponíveis para participar dos testes, o avaliador pode convidar para o teste-piloto uma pessoa com perfil um pouco diferente do desejado. Para o teste-piloto, não há muito problema em convidar um colega de trabalho, um amigo, ou uma pessoa mais próxima do avaliador, pois os dados coletados serão descartados. Entretanto, esse participante não deve ter se envolvido no planejamento e na preparação da avaliação.

Com o planejamento da avaliação concluído, o avaliador recruta os participantes para as sessões de avaliação. Para **recrutar participantes** com os perfis especificados, o avaliador pode utilizar questionários ou entrevistas curtas a fim de conferir se uma pessoa possui o perfil desejado. Ao convidar os participantes selecionados, o avaliador deve esclarecer brevemente quais são os objetivos da avaliação, como e onde ela será realizada, e quanto tempo será requerido do participante. Desde o primeiro contato com o participante, o avaliador deve estar aberto e disposto a esclarecer suas eventuais dúvidas sobre a avaliação. Confirmada a disponibilidade e o interesse do participante, o avaliador deve agendar com ele a data, o horário e o local para a realização da avaliação.

Sempre que possível, o avaliador deve evitar selecionar seus conhecidos, amigos, alunos, subordinados ou pessoas com as quais possua alguma relação próxima, para evitar que a relação entre ele e o participante influencie ou contamine os dados coletados. Por exemplo, um amigo e um aluno podem não relatar problemas que perceberam na interface por se sentirem criticando o trabalho do avaliador amigo ou professor. De qualquer modo, o avaliador pode atenuar a interferência das relações entre eles esclarecendo que o objetivo da avaliação é encontrar problemas na interface e que a opinião dos participantes é importante para melhorá-la.

### 11.7.3 Coleta de Dados

A **coleta de dados** deve ocorrer conforme o planejamento realizado e o método de avaliação selecionado. No caso de métodos de avaliação por inspeção, essa atividade envolve apenas os avaliadores, que utilizam o material preparado e seguem o procedimento prescrito pelo método selecionado. Os avaliadores examinam a interface para identificar discrepâncias com um padrão ou para tentar *prever* as experiências de uso que os usuários terão com a solução de IHC avaliada (veja Seção 12.2).

Em avaliações por investigação e por observação, essa atividade costuma ter por objetivo registrar as experiências vivenciadas pelos usuários durante a interação com o sistema ou protótipo sendo avaliado. Os métodos empíricos contam com a participação dos usuários para *relatar* experiências de uso vivenciadas ou *permitir a observação* de experiências reais de uso com a solução de IHC avaliada.

Ao receber os participantes, o avaliador deve ser cordial e deixá-los bem à vontade. Ele costuma estabelecer uma conversa de aquecimento ou “quebra-gelo” com os participantes, sobre tópicos gerais como: o clima, notícias recentes e como foi o trajeto até o local da avaliação. Se a avaliação ocorrer em laboratório, o avaliador deve apresentar todo o ambiente de teste aos participantes, inclusive a sala de observação. Sempre que possível, o avaliador deve oferecer água, café, oportunidade para ir ao toalete ou algo mais de que os participantes precisem. Tudo isso tem por objetivo dar oportunidade e tempo para o participante se acostumar com o ambiente, reduzindo sua ansiedade antes de iniciar a sessão de coleta de dados.

Depois dessa conversa inicial, o avaliador deve explicar ao participante os objetivos do estudo, o sistema de interesse, o procedimento da avaliação (por exemplo, apresentar uma visão geral das atividades que o participante é convidado a realizar) e os cuidados éticos sendo tomados (veja Seção 7.4). Se surgir alguma dúvida, ela deve ser sanada imediatamente pelo avaliador. Depois, o avaliador entrega o termo de consentimento em duas vias, já assinadas por ele próprio, e aguarda enquanto o participante lê e assina o termo, caso aceite participar da avaliação. Uma via do termo de consentimento fica com o avaliador e a outra com o participante.

**Em seguida, o participante responde** o questionário pré-teste. O avaliador pode tornar esse momento mais natural se ele ler as perguntas do questionário para o participante e registrar as respostas fornecidas, como numa entrevista estruturada. Isso pode ajudar o participante a se sentir mais à vontade durante a avaliação.

Respondido o questionário, é iniciada a sessão de observação. Nesse momento, o software e os equipamentos que registram os dados da interação devem ser ativados. O avaliador apresenta o sistema a ser avaliado e pode permitir que o participante explore-o livremente por alguns minutos, caso seja o seu primeiro contato com o sistema. Depois da exploração livre, o avaliador entrega ao participante as instruções e os cenários das tarefas a serem realizadas. Esclarecidas as eventuais dúvidas, o participante passa a realizar as tarefas solicitadas.

É recomendável que pelo menos dois avaliadores trabalhem na coleta de dados: um para acompanhar o participante mais de perto na sala de uso do sistema e outro mais distante da sala de observação. Enquanto o participante utiliza o sistema, os avaliadores devem anotar qualquer acontecimento que possa ser relevante para a interpretação dos dados coletados e sobre eventuais dúvidas a serem esclarecidas na entrevista pós-teste. Eles devem evitar interferir na atividade do usuário. O participante não deve ser interrompido nem questionado enquanto realiza as tarefas solicitadas.

Se for adequado, o avaliador pode pedir que o participante relate em voz alta o que ele está pensando e fazendo durante a interação. Essa técnica é conhecida como *think aloud* (“pensar em voz alta”) (Ericsson e Simon, 1993). Apesar de ser uma técnica útil e relativamente simples para o avaliador ter acesso ao que se passa na mente do participante, ela depende muito do participante. Certas pessoas podem se distrair e parar de falar enquanto realizam alguma atividade. Outras podem gastar mais tempo e esforço relatando o que estão pensando e fazendo, do que realmente executando as tarefas solicitadas. Se não for aplicada de

forma adequada, essa técnica pode interferir nos resultados da avaliação, como, por exemplo, no número de erros cometidos, pois o usuário pode pensar mais antes de fazer, ou no tempo de execução das tarefas, pois o participante pode gastar mais tempo pensando e falando. Portanto, a técnica de *think aloud* deve ser utilizada cuidadosamente.

Caso o participante já tenha gasto muito tempo interagindo sem conseguir concluir uma tarefa, demonstre não ter condições de concluir a tarefa, ou passe por uma situação de constrangimento ou emoções desagradáveis, os avaliadores podem intervir na interação, sugerindo ao participante que passe para a próxima tarefa solicitada, ou até mesmo abandone as tarefas restantes. Na transição entre tarefas, se for o caso, o avaliador pode interagir com o sistema a fim de prepará-lo para a execução da próxima tarefa. Além disso, o participante pode pedir para fazer uma pausa, sempre que quiser.

Terminada a sessão de interação, os avaliadores devem conduzir a entrevista pós-teste. Essa entrevista é uma oportunidade para coletar a opinião do participante sobre a experiência de uso que acabou de vivenciar e esclarecer eventuais dúvidas sobre seu comportamento e suas intenções, percepções e interpretações durante a execução das tarefas.

### 11.7.4 Interpretação

Na atividade de **interpretação**, o avaliador analisa o material registrado para atribuir significado aos dados coletados. A interpretação do avaliador deve ser orientada pelo método de avaliação selecionado e pelo que foi definido durante a atividade de preparação da avaliação. Cada método de avaliação costuma apontar os focos de análise (*i.e.*, quais dados devem ser analisados sob quais perspectivas de análise) e os tipos de interpretações mais frequentes. Por exemplo, o método de avaliação heurística enfatiza a análise de um conjunto de heurísticas, enquanto o método de avaliação de comunicabilidade investiga problemas na recepção da metamensagem associados a etiquetas de rupturas de comunicação. Em geral, o avaliador se concentra inicialmente na interpretação dos dados coletados de cada participante individualmente, buscando respostas às questões específicas definidas no planejamento da avaliação. Esse tipo de análise também é conhecido como análise *intrassujeito* ou *intraparticipante* (Nicolaci-da Costa, 1994; Nicolaci-da Costa et al., 2004).

A interpretação ou análise dos dados coletados pode ser feita de forma *automática* ou *manual*, dependendo do tipo de cada dado. Alguns aspectos de uma solução de IHC podem ser avaliados automaticamente por um programa computacional, de forma mais rápida e sistemática do que um ser humano poderia realizar manualmente. A análise automática dos dados tem sido utilizada para avaliar a navegação por sites na Web, verificando, por exemplo, a existência de “links quebrados” e alguns critérios de acessibilidade (*e.g.*, imagens que não possuem descrição textual associada). O DaSilva<sup>1</sup> é um dos primeiros programas brasileiros que avaliam a acessibilidade de websites. Ele analisa o código HTML do endereço indicado para detectar partes da interface que violem regras de acessibilidade.

Também é possível verificar automaticamente se o uso do sistema avaliado ocorre conforme o esperado. Para isso, na etapa de preparação o avaliador representa em algum modelo os caminhos de interação oferecidos pelo designer para o usuário alcançar seus objetivos (por exemplo, através de um modelo de tarefas ou de interação). Na etapa de coleta de dados, um programa monitora e registra a interação do usuário com o sistema enquanto ele tenta atingir os objetivos desejados. Desse modo, o avaliador pode comparar automaticamente o que ele esperava que ocorresse e o que de fato ocorre durante a interação. Dentre outros resultados, essa comparação permite calcular quantas vezes os usuários conseguiram ou não alcançar seus objetivos da forma prevista pelo designer, bem como identificar em que partes da interação ocorreram problemas que dificultaram ou impediram os usuários de alcançarem seus objetivos (Lecroff e Paterno, 1998). Essa é uma forma de avaliação baseada em modelos e registros de interação.

<sup>1</sup><http://www.acessibilidadebrasil.org.br/joomla/noticias/206-utilize-o-primeiro-programa-nacional-para-a-avaliacao-da-acessibilidade-de-sites>

Os critérios de qualidade de uso que podem ser analisados automaticamente por programas computacionais ainda são muito limitados, pois exigem uma definição a priori de regras definindo o que é “bom” e “ruim”, e o que é “certo” e “errado”. A análise realizada por um avaliador humano ainda é fundamental para verificar a qualidade de uso, porque é difícil codificar num programa toda a visão que o avaliador pode adquirir sobre domínio, usuário, atividades e contexto, bem como sua capacidade de análise, principalmente diante de situações imprevistas. No entanto, para aspectos pontuais da avaliação, as análises automáticas podem agilizar parte do processo.

Grupo 29

### 11.7.5 Consolidação e Relato dos Resultados

Uma vez concluída a interpretação individual dos dados coletados, seja das previsões dos avaliadores ou das observações das experiências de uso dos participantes, os resultados individuais são **consolidados** e analisados em conjunto, em uma análise denominada de intersujeito ou interparticipante (Nicolaci-da Costa, 1994; Nicolaci-da Costa et al., 2004). Nessa atividade, os avaliadores buscam recorrências nos resultados de acordo com o método selecionado. As **recorrências** são importantes porque, ao expressarem resultados comuns a vários participantes de um grupo, permitem fazer uma distinção entre características representativas do grupo e as idiossincrasias de participantes individuais.

Na consolidação dos resultados, os avaliadores devem novamente endereçar as questões que motivaram o estudo, buscando respondê-las ou justificar por que alguma resposta não foi encontrada. Mesmo no caso de avaliações empíricas, a **generalização** dos resultados exige muito cuidado. Eles sempre são fortemente influenciados pelo ambiente de avaliação e pelas características, preferências, interesses e necessidades dos participantes individuais.

Os métodos de avaliação podem apresentar diferentes resultados. Por exemplo, quando o objetivo é identificar problemas em soluções de IHC, diferentes métodos podem revelar tipos diversos de problemas que prejudiquem a qualidade de uso, seja por causa de diferenças na coleta de dados ou na sua análise. Outros fatores também influenciam os resultados da avaliação: o conhecimento e a experiência dos avaliadores, o tempo disponível para a avaliação, a quantidade e qualidade dos dados disponíveis, e assim por diante.

Grupo 30

**Os resultados de uma avaliação** de IHC normalmente indicam tendências de problemas, e não uma certeza de que eles vão ocorrer durante o uso do sistema. De modo semelhante, caso não sejam encontrados problemas durante a avaliação, também não podemos afirmar categoricamente que o sistema tenha alta qualidade de uso, apenas que um estudo não revelou problemas num determinado escopo do sistema que foi avaliado com base em um certo planejamento.

Finalmente, os avaliadores devem **relatar os resultados** consolidados, que costumam incluir:

- os objetivos e escopo da avaliação;
- a forma como a avaliação foi realizada (método de avaliação empregado);
- o número e o perfil de usuários e avaliadores que participaram da avaliação;
- um sumário dos dados coletados, incluindo tabelas e gráficos;
- um relato da interpretação e análise dos dados;
- uma lista dos problemas encontrados;
- um planejamento para o reprojeto do sistema.

Grupo 31

## 11.8 O Framework DECIDE

Preece et al. (2002) propõem um *framework* chamado DECIDE para orientar o planejamento, a execução e

a análise de uma avaliação de IHC. As atividades do *framework* são interligadas e executadas iterativamente, à medida que o avaliador articula os objetivos da avaliação, os dados e recursos disponíveis. Então, quando o avaliador descobre uma necessidade de modificar os rumos da avaliação por algum motivo, as demais atividades são afetadas. Por exemplo, se o avaliador não conseguir permissão para visitar o ambiente de uso de um sistema, ele não pode aplicar um método de avaliação que coleta dados sobre o uso do sistema em contexto. Nesse caso, provavelmente seus objetivos precisarão ser revistos. As atividades do *framework* DECIDE são descritas a seguir.

- D Determinar os objetivos da avaliação de IHC. O avaliador deve determinar os objetivos gerais da avaliação e identificar por que e para quem tais objetivos são importantes. O restante do planejamento da avaliação, sua execução e a apresentação dos resultados serão orientados por esses objetivos.
- E Explorar perguntas a serem respondidas com a avaliação. Para cada objetivo definido, o avaliador deve elaborar perguntas específicas a serem respondidas durante avaliação. Essas perguntas são responsáveis por operacionalizar a investigação e o julgamento de valor a serem realizados. Elas devem considerar o perfil dos usuários-alvo e suas atividades.
- C Escolher (*Choose*) os métodos de avaliação a serem utilizados. O avaliador deve escolher os métodos mais adequados para responder as perguntas e atingir os objetivos esperados, considerando também o prazo, o orçamento, os equipamentos disponíveis e o grau de conhecimento e experiência dos avaliadores.

**Grupo 32** **I** Identificar e administrar as questões práticas da avaliação. Existem muitas questões práticas envolvidas numa avaliação de IHC, como, por exemplo, o recrutamento dos usuários que participarão da avaliação, a preparação e o uso dos equipamentos necessários, os prazos e o orçamento disponíveis, além da mão-de-obra necessária para conduzir a avaliação.

- D Decidir como lidar com as questões éticas. Sempre que usuários são envolvidos numa avaliação, o avaliador deve tomar os cuidados éticos necessários (veja Seção 7.4). Os participantes da avaliação devem ser respeitados e não podem ser prejudicados direta ou indiretamente, nem durante os experimentos, nem após a divulgação dos resultados da avaliação.
- E Avaliar (*Evaluate*), interpretar e apresentar os dados. O avaliador precisa estar atento a alguns aspectos da avaliação realizada antes de tirar conclusões e divulgar resultados. Ele deve considerar: o grau de confiabilidade dos dados (*i.e.*, semelhança dos resultados obtidos quando emprega mais de uma vez o mesmo método de avaliação nas mesmas circunstâncias; a validade interna do estudo (*i.e.*, se o método de avaliação mede o que deveria medir, se o faz com rigor e evita que os dados sejam distorcidos); a validade externa do estudo (*i.e.*, até que ponto os resultados podem ser generalizados ou transferidos a um outro contexto semelhante); e a validade ecológica do estudo (*i.e.*, o quanto os materiais, métodos e ambiente de estudo se assemelham à situação real investigada).

## Atividades

1. *Por que avaliar o uso de software?* Imagine que você seja apresentado a um produtor de software que lhe conta sobre seus planos para produzir uma nova versão de um sistema comercial. Quais argumentos você elaboraria para convencê-lo a realizar uma avaliação de IHC?
2. *Planejamento da avaliação de IHC.* Escolha um software de sua preferência e defina dois objetivos de avaliação (por exemplo, dois objetivos citados na Tabela 11.1). Planeje duas avaliações de IHC do software escolhido, uma para cada objetivo definido. Em cada planejamento, realize cada tarefa de preparação da avaliação e relate sua execução. Compare os dois planejamentos.