

# Desafios para Testes de Usuários em Sistemas Colaborativos - Lições de um Estudo de Caso

Raquel Oliveira Prates

Departamento de Ciência da Computação – UFMG

Av. Antônio Carlos 6627, ICEx, sala 4010, Pampulha,  
Belo Horizonte - MG, CEP 31270-010

rprates@dcc.ufmg.br

Alberto Barbosa Raposo

TeCGraf - PUC-Rio

Rua Marquês de São Vicente, 225, Gávea, Rio de  
Janeiro – RJ, CEP: 22453-900

abraposo@tecgraf.puc-rio.br

## ABSTRACT

Collaborative systems evaluation is still an open issue for researchers working on the fields of HCI and CSCW. In this paper we present the challenges experienced by evaluators in a real system evaluation. These challenges and the discussion about them contribute to the research in the field, as well as to the planning of evaluations involving users for collaborative systems.

## Categories and Subject Descriptors

D.2.2: User Interfaces; H.5.3 Group and Organization Interfaces: Computer supported cooperative work, Evaluation/methodology

## Keywords

Collaborative systems, evaluation, user testing.

## 1. INTRODUÇÃO

Em 1988, Grudin discutiu a importância de se avaliar sistemas colaborativos e os desafios envolvidos para tal [3]. Quase duas décadas depois de a questão ter sido levantada, ela ainda desafia a comunidade que atua na área de sistemas colaborativos. Muitas pesquisas têm tentado entender como as avaliações têm sido feitas [4], estudar os desafios envolvidos e propor novos modelos ou métodos [1,5]. No entanto, apesar das pesquisas na área, ainda não existe consenso sobre o que avaliar, que método utilizar e quando avaliar. Ainda é necessário gerar conhecimento sobre os problemas vivenciados na avaliação de sistemas colaborativos, para que se possa oferecer aos designers informações que possam apoiar suas decisões relativas à avaliação.

O alto custo da avaliação com usuários para sistemas colaborativos tem encorajado a investigação sobre métodos de avaliação de mais baixo custo [1,5]. No entanto, normalmente estas propostas são de métodos de inspeção que não envolvem usuários. Em um trabalho comparando os resultados da inspeção x avaliação com os usuários Steves e seus colegas mostram a importância de se fazer testes com usuários também em sistemas colaborativos [7]. Embora isto não seja novidade na área de IHC e Grudin já tenha mostrado que conhecimento sobre design de sistemas mono-usuário e bom senso não são suficientes para se

tomar boas decisões relativas a sistemas colaborativos, pouco se caminhou na direção de auxiliar avaliadores nas decisões e planejamento de um teste com usuários de sistemas colaborativos.

O trabalho apresentado neste artigo é um passo nesta direção. Neste trabalho descrevemos um estudo de caso em que se fez o planejamento e execução parcial da avaliação com usuários de um sistema colaborativo de apoio à comunicação síncrona. Identificamos os desafios vivenciados pelos avaliadores discutindo os aspectos envolvidos e ilustrando com o estudo de caso. Estes desafios e a discussão apresentada podem auxiliar avaliadores que pretendam avaliar sistemas colaborativos nas suas considerações e decisões.

## 2. ESTUDO DE CASO

### 2.1 CSVTool

O CSVTool (*Collaboration Supported by Video Tool*) é uma ferramenta desenvolvida para uma grande companhia de exploração de Petróleo [6]. Como a companhia tem unidades tanto administrativas quanto de produção espalhadas por diversas regiões do Brasil, a comunicação entre estas unidades se faz necessária. O CSVTool foi projetado em parceria com o centro de pesquisas desta companhia para atender às suas necessidades de comunicação e também a requisitos tecnológicos, tais como independência de plataforma. Assim, o principal objetivo do sistema é oferecer aos membros de um grupo recursos de comunicação, como canais de áudio, vídeo e texto (Figura 1).

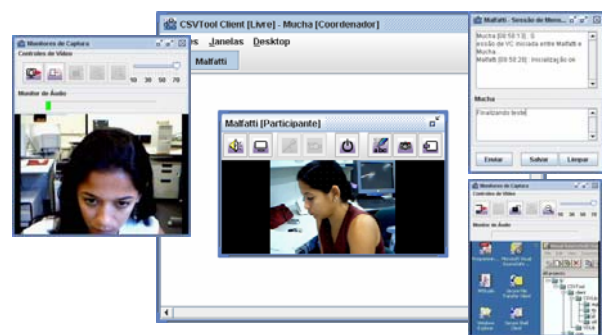


Figura 1 – Tela principal do CSVTool

O projeto do CSVTool foi feito de forma que ele possa tanto ser utilizado como um sistema independente de video-conferência quanto integrado a outros sistemas colaborativos já em uso pelos funcionários desta companhia. A maior parte dos funcionários atualmente utiliza sistemas mono-usuário no seu cotidiano na empresa, mas não têm costume de usar sistemas colaborativos. Os funcionários normalmente se comunicam pessoalmente ou por

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference '04, Month 1–2, 2004, City, State, Country.  
Copyright 2004 ACM 1-58113-000-0/00/0004...\$5.00.

telefones. Alguns utilizam sistemas de mediação de comunicação simples como sistemas de mensagem instantânea. Desta forma um dos objetivos do CSVTool era ser fácil de utilizar e inicializar e facilmente integrado a outras aplicações. Embora o projeto do CSVTool já tenha uma versão piloto que está sendo utilizada por técnicos em geociências tanto na sua versão independente, quanto integrada a uma aplicação de modelagem geológica, não havia sido feita nenhuma avaliação da sua interface. Desta forma, foi planejada uma avaliação para se apreciar a qualidade da interface da versão independente do CSVTool.

O projeto da avaliação da interface do CSVTool previa duas etapas. Na primeira, especialistas na área de IHC fariam uma inspeção informal do sistema para definirem o teste com usuários. Na segunda, seriam feitos os testes com os usuários.

## 2.2 Avaliação Executada

A primeira etapa da avaliação consistiu na inspeção do CSVTool por avaliadores experientes na área de IHC. A avaliação foi informal, pois embora tenha sido detalhada, não foi guiada por nenhum método específico, mas apenas pela experiência dos avaliadores. Inicialmente a idéia era que cada avaliador examinaria o CSVTool independentemente e depois discutiriam os problemas identificados para concluir (1) que aspectos deveriam ser avaliados com os usuários; (2) como deveria ser o teste com os usuários. No entanto, a avaliação independente se mostrou inviável, pois sendo o CSVTool uma ferramenta de comunicação síncrona, era praticamente impossível simular uma conferência para se fazer a inspeção. Desta forma, a inspeção foi feita pelos avaliadores juntos. A medida que se comunicavam pelo sistema, faziam a inspeção explorando a interface, e identificando potenciais problemas. A partir da inspeção foi gerada uma lista das questões que poderiam causar dificuldades no seu uso. Os principais problemas identificados foram:

1. Convite à participação: quando o usuário aceitava o convite, ele automaticamente começava a participar da comunicação. No entanto, quando o negava, quem o havia convidado não tinha nenhum feedback; o usuário simplesmente não aparecia. Além disso, o convite não expirava; a qualquer momento posterior o usuário que negou o convite poderia decidir participar da comunicação.
2. Canais de comunicação por áudio/vídeo: Os usuários tinham um grande controle dos canais de comunicação de vídeo e áudio, podendo desligar separadamente o envio ou recepção deste de/para qualquer outro usuário específico participando da sessão. O custo deste fino controle era a dificuldade de percepção do status da comunicação do usuário com os demais participantes da sessão. O status com cada participante era comunicado por 4 diferentes botões na interface (relativos ao envio/recebimento de áudio/vídeo), sendo que cada botão podia assumir 4 estados distintos.

Os avaliadores identificaram que estes problemas, além de dificultar a interação, poderiam forçar um comportamento do usuário não condizente com as normas sociais e culturais vigentes, gerando assim problemas de natureza social relacionados à polidez e privacidade. Por exemplo, quando o usuário estava ocupado no momento de um convite e não o aceita. Se posteriormente ele ficou disponível, ele não tem como avisar (pelo sistema) a quem o havia convidado que já está liberado, ou mesmo enviar um convite à outra pessoa. Ao tentar convidar a

pessoa para uma conversa, se a sessão em que o convite fora feito a ele ainda estivesse ativa, ele seria automaticamente conectado à conversa. Esta entrada repentina na conversa poderia causar interrupções ou mesmo ser inconveniente. Outro exemplo poderia ser causado pelo fato de a interface mostrar o status de comunicação de cada participante com cada outro, sem oferecer uma visão geral. Assim, embora na cultura brasileira seja rude deixar uma pessoa falando sozinha, no CSVTool esta situação poderia ocorrer, sem que os participantes (além do falante) se dessem conta. Por exemplo, se alguém estivesse falando com vários outros pelo CSVTool, e por algum motivo todos os ouvintes interrompessem a recepção de áudio sem saber que os demais tinha feito o mesmo, o falante seria deixado falando sozinho. O falante poderia, por exemplo, interpretar esta ação dos demais participantes como sendo “Não estamos interessados no que você está dizendo”.

Ao conversar com o gerente do projeto sobre as conclusões da inspeção, ele disse que realmente não tinham considerado que estas situações poderiam levar a desconfortos ou problemas sociais. A intenção de design ao oferecer um controle tão fino dos canais de comunicação era permitir aos usuários um melhor controle sobre os recursos disponíveis. Assim, considerando que alguns usuários poderiam ter diferentes qualidades de conexão disponíveis, este controle permitia a cada um decidir que canais seriam relevantes de ficarem abertos ou não. Por exemplo, no caso de a conversa ter um falante principal, o usuário (por falta de banda) poderia interromper o envio/recebimento de vídeo com os demais participantes, mantendo apenas o vídeo do falante aberto.

Dado que os avaliadores não foram capazes de perceber que era esta a intenção de design, um dos pontos relevantes da avaliação em questão deveria ser se o usuário conseguiria entender esta intenção do design. Este entendimento poderia minimizar alguns potenciais efeitos negativos de natureza social. Por exemplo, no caso de todos os ouvintes interromperem o recebimento do áudio, o falante talvez estivesse mais propenso a interpretar isto como “Parece que a conexão não está muito boa” ao invés de interpretar como desinteresse dos demais.

Com base nos potenciais problemas identificados pelos avaliadores e também na reunião com o gerente de projeto, foi decidido como seria a avaliação com os usuários. Os principais objetivos do teste com os usuários foram definidos como:

1. Observar se conseguiriam manter a conversa fluente;
2. Observar se conseguiriam se manter a par do status de sua comunicação com os demais membros de um grupo pequeno;
3. Avaliar se conseguiriam entender a intenção de design em relação ao controles dos canais de comunicação;
4. Observar sua reação em situações sociais que geravam potenciais problemas de privacidade e polidez;

Para se conseguir fazer a apreciação destes pontos com os usuários, foi definido que seria feito um teste com grupos de 3 usuários em um ambiente controlado, onde os usuários estariam distribuídos, tendo apenas o CSVTool como meio de comunicação. Foi preparado um cenário no qual um participante seria entrevistado pelos outros dois sobre um tema sendo amplamente discutido pela sociedade na época. Para conseguir simular as situações desejadas, foi feito um script distinto para cada usuário, e definido que cada um seria acompanhado por um

avaliador, para poder observar suas reações e também garantir que seguiriam o script (e.g., lembrando-o da sua vez de falar).

O cenário de teste previa que o participante entrevistadoalaria livremente sobre o tema, respondendo a perguntas preparadas para os demais participantes. No decorrer da entrevista os outros participantes fariam mudanças no status de sua comunicação com o entrevistado, sendo que em um dado momento ele era deixado falando sozinho (sem que ninguém o estivesse escutando). Além disso, um quarto participante (membro da equipe de design) era convidado a participar da conferência, mas só entrava na conferência perto do seu fim. Após o teste, estava prevista uma entrevista coletiva contando com a participação de todos os avaliadores e participantes do teste.

O teste estava previsto para ser executado com 5 grupos de pessoas com perfis similares aos dos profissionais da companhia, já que a participação dos funcionários da empresa não seria viável. Para avaliar o teste planejado foi feito um teste piloto.

O teste piloto deixou claro um entendimento errôneo que os avaliadores tiveram da forma como funcionava o convite. Os avaliadores haviam entendido que o convite era feito para a pessoa participar de uma conferência, e que uma vez aceito o convite, este participante estaria conectado a todos os demais participantes presentes. No entanto, o convite é individual, ou seja, aceitando o convite o participante se conecta apenas a quem o convidou e não a todos os outros. Durante o teste piloto, os avaliadores rapidamente identificaram o problema, e deram instruções para superá-lo. No entanto, o script teve que ser mudado e um novo teste piloto projetado. O segundo teste piloto correu bem e o plano de avaliação foi considerado pronto para a avaliação com os usuários. Porém, o alto custo do teste e de seu planejamento, aliados à obtenção de alguns indicadores de alguns problemas significativos da interface do CSVTool, resultaram na decisão de não prosseguir com os testes naquele momento.

### 3. DESAFIOS IDENTIFICADOS

Nesta seção apresentamos os desafios vivenciados durante a avaliação do CSVTool. O objetivo de identificá-los é justamente permitir a outros avaliadores planejando testes com usuários de sistemas colaborativos refletir sobre eles durante a etapa de planejamento evitando alguns dos problemas vivenciados. Vale ressaltar que a equipe de avaliação contava com especialistas com grande experiência em avaliação de interação. Os especialistas tinham também experiência em avaliação de sistemas colaborativos, já tendo planejado e executado tanto avaliações por inspeção, quanto testes com usuários destes sistemas.

#### **Desafio 1: Determinar o número necessário de avaliadores para a avaliação**

As pesquisas na área de avaliação de sistemas colaborativos não discutem esta questão. Na maior parte dos casos se descreve o que foi feito [4]; ou em casos de propostas de métodos [1,5] adota-se uma solução que funciona em casos de sistemas mono-usuários sem uma discussão maior sobre o tema. No entanto, no caso do CSVTool soluções para sistemas mono-usuários não foram facilmente traduzidas para a avaliação necessária, nem para a avaliação de inspeção, nem para o teste com usuários.

A inspeção feita no CSVTool foi propositadamente informal, pois seu objetivo era o planejamento da avaliação com os usuários. Mesmo assim, ela apontou para algumas considerações a serem

feitas no caso de se planejar uma avaliação por inspeção em um sistema colaborativo síncrono. (No assíncrono é provável que o avaliador consiga simular a colaboração assumindo diferentes papéis a cada momento). Como mostramos, a inspeção do CSVTool não era possível de ser feita individualmente, portanto, ela teve que ser feita por dois avaliadores. Mesmo assim, alguns pontos sobre a interação não foram compreendidos pelos avaliadores, como ficou claro posteriormente, durante o primeiro teste piloto. Para que se tivesse sido possível identificar o problema da diferença do convite para a conversa a dois com relação ao convite para a conferência, teriam sido necessárias no mínimo três pessoas na conferência. A maioria dos métodos de inspeção recomenda que se tenha de 3 a 5 avaliadores produzindo avaliações independentes para se identificar um número significativo de problemas. Neste caso, seriam necessárias no mínimo 3 sessões de inspeção, cada uma com 3 pessoas, ou seja no mínimo 9 pessoas envolvidas na avaliação. Para se abaixar o custo poderia se ter apenas um avaliador por sessão, e outras 2 pessoas colaborando para a interação (e.g., poderiam ser da própria equipe de design).

No caso do CSVTool, o cenário foi projetado com o objetivo de simular algumas situações de uso, para garantir que os pontos de interesse pudessem ser observados. No entanto, neste caso, o teste exigia um alto nível de coordenação, o que significava ter um avaliador ao lado de cada participante. Este custo era necessário, pois caso um dos participantes deixasse de alterar o status de sua comunicação, ou a entrevista fosse interrompida (e.g., um usuário esquecer de fazer a pergunta na sua vez) não seria possível observar os pontos desejados, e talvez o teste tivesse que ser desprezado. Ademais, para se observar a reação das pessoas, deve-se estar presente ou filmá-las, o que neste caso exigiria mais equipamentos e, posteriormente, tempo de análise.

Embora em tempo de planejamento a necessidade dos avaliadores tivesse sido prevista, mesmo assim ela se mostrou inviável. Assim, uma sugestão teria sido ter feito o cenário para observar um usuário por vez, contando com 2 outros participantes que pudessem seguir o script, garantindo as situações desejadas para o usuário. Os participantes poderiam ser da própria equipe de design, ou colaboradores experientes em tecnologias de comunicação. Isto permitiria que o teste fosse feito com apenas um avaliador por vez e diminuiria também o número de voluntários necessários (desafio discutido a seguir).

#### **Desafio 2: Dificuldade de se conseguir o número necessário de participantes para o teste**

Para se fazer um teste com usuários de natureza qualitativa normalmente se recomenda ter de 5 a 8 participantes. No entanto, no caso do CSVTool isto representava 5 grupos de 3 participantes, ou seja, 15 participantes. No caso do CSVTool, em que os usuários eram profissionais da indústria, não era viável contar com a participação dos usuários no teste. Assim, a alternativa era conseguir voluntários (já que a legislação brasileira não permite que a participação no teste seja remunerada [2]) com o perfil similar aos dos usuários. As características principais identificadas eram pessoas que não fossem da área tecnológica, mas usassem computador na execução de suas tarefas. Além da dificuldade de se conseguir pessoas com este perfil dispostas a doar seu tempo para a avaliação, havia ainda um complicador, que era coordenar a disponibilidade destes voluntários para se juntar os grupos de 3. Para o teste piloto os voluntários eram do

próprio laboratório (ou seja, não atendiam a característica de não serem da área tecnológica), mas não conheciam o CSVTool.

### **Desafio 3: Dificuldade de se avaliar durante o teste questões de natureza social e cultural geradas pela tecnologia**

Observar durante testes com usuários em laboratórios questões relacionadas com o impacto da tecnologia no grupo ou organização, e seu efeito nas relações sociais é de fato impossível, tanto pela natureza de simulação do teste, quanto pelo tempo de uso da ferramenta. Assim, para se observar estas questões seria realmente necessário acompanhar o uso da ferramenta no seu contexto real. De todo jeito, durante o teste é possível colher alguns indicadores que possam ser úteis na avaliação do potencial impacto do sistema, ou mesmo no planejamento de avaliações mais longas acompanhando o uso do sistema. Mesmo a identificação de quais seriam estes indicadores e a definição de cenários que permita observá-los nem sempre é fácil, ou pode trazer outros custos associados.

No caso do CSVTool o objetivo era justamente identificar se teríamos indicadores de como o usuário entenderia a função proposta do controle fino dos canais de comunicação e o uso que os outros usuários faziam dele (relacionando-os com falta de polidez e gerando um impacto social negativo ou relacionando-os com problemas tecnológicos). Na avaliação do CSVTool os indicadores pareceram apropriados durante os testes piloto. No entanto, estes indicadores eram a principal causa de um cenário com necessidade de um alto nível de coordenação.

### **Desafio 4: Identificação de problemas básicos que justificam a interrupção da avaliação**

Em um estudo de caso de avaliação de sistemas colaborativos Twidale e colegas [8] definiram um problema básico (*scale-up problem*) como sendo um problema que impedia o uso pretendido do sistema em algum aspecto, impedindo que se avaliasse o sistema em relação a eles. Assim, a identificação deste problema justifica interromper a avaliação, acertar o sistema, para então fazer nova avaliação do sistema.

No CSVTool o grande número de botões e estados necessários de serem analisados para identificar o status da comunicação dos participantes com os demais impediu que os usuários se mantivessem informados sobre isso. Assim, durante o teste piloto nenhum usuário foi capaz de identificar as situações simuladas (e.g., o entrevistado ter ficado falando sozinho). Como as situações de potencial impacto social negativo não foram sequer identificadas, não foi possível observar a reação dos usuários a elas. Os indicadores sobre elas foram obtidos durante a entrevista, ao informar os usuários de que a situação ocorrera. Embora, neste caso isto até possa parecer uma consequência positiva, na verdade a falta de visão sobre o status da comunicação potencialmente pode dificultar a comunicação e causar outros problemas, como acreditar que todos estão cientes de um fato apresentado em uma conferência, sem se perceber que as pessoas não podiam ouvir o que estava sendo dito. Desta forma, considerou-se que este problema era básico e justificava ser tratado antes de se avaliar os pontos relativos ao entendimento do controle e impactos sociais.

A dificuldade neste caso muitas vezes é identificar se é possível se caracterizar um problema como básico antes mesmo de se fazer os testes com os usuários. No caso do CSVTool esta decisão foi tomada, uma vez que os especialistas tinham-no identificado e nos

dois testes pilotos feitos nenhum dos usuários conseguiu perceber as interrupções feitas nos canais de comunicação.

## **4. CONSIDERAÇÕES FINAIS**

Neste artigo levantamos a partir de um projeto real de avaliação desafios e decisões a serem considerados durante o planejamento de uma avaliação de usuários de sistemas colaborativos. Ilustramos com a experiência vivenciada as considerações envolvidas em cada um dos desafios. Assim, contribuimos para a geração de conhecimento sobre avaliações com usuários de sistemas colaborativos. Este conhecimento pode auxiliar tanto avaliadores durante a etapa de decisões relativas à avaliação a ser feita de um sistema colaborativo, quanto a pesquisa na área.

Finalmente, os próximos passos na área deste trabalho envolvem tanto planos referentes ao desenvolvimento do CSVTool, quanto à pesquisa na área. Em relação ao CSVTool, com base nos problemas encontrados durante o planejamento da avaliação e testes piloto, a interface está sendo reprojeta. Assim que for concluído o reprojeto, novas avaliações serão executadas, levando em conta o conhecimento adquirido na experiência relatada, e explorando algumas das alternativas discutidas. Em relação à pesquisa, o objetivo é coletar mais conhecimento sobre aspectos que influenciam e dificultam a avaliação de sistemas colaborativos e tentar organizar este conhecimento na forma de um esquema ou framework que possa guiar as decisões envolvidas no planejamento deste tipo de avaliação.

## **5. AGRADECIMENTOS**

O Tecgraf é prioritariamente financiado pela Petrobras. Raquel Prates agradece o CNPq pelo financiamento no. 475992/2003-2.

## **6. REFERÊNCIAS**

- [1] Baker, K., Greenberg, S. and Gutwin, S. (2001) "Heuristic evaluation of groupware based on the mechanics of collaboration", Proceedings of EHCI'01. LNCS, V. 2254, 123-139, Springer-Verlag.
- [2] CNS - Conselho Nacional de Saúde. (1996) Resolução número 196/96 sobre pesquisas envolvendo seres humanos. <http://www.datasus.gov.br/conselho/reso196/RES19696.htm>.
- [3] Grudin, J. (1988). Why groupware applications fail: Problems in design and evaluation. Office: Technology and People, 4(3), 245-264.
- [4] Pinelle, D. and Gutwin, C. (2000) A review of groupware evaluations, Proceedings of 9th IEEE WETICE 2000, USA.
- [5] Pinelle, D. and Gutwin, C. (2002) Groupware walkthrough: adding context to groupware usability evaluation, Proceedings of CHI 2002, p.455-462.
- [6] Pozzer, C. T., Lima, L. S., Raposo, A. B., Vieira, C. J. G., (2005) A Multi-user Videoconference-based Collaboration Tool: Design and Implementation Issues. Proceedings of CSCWD 2005, p. 547-552, U.K.
- [7] Steves, M. P, Morse, E., Gutwin, G., Greenberg, S. "A Comparison of Usage Evaluation and Inspection Methods for Assessing Groupware Usability". Group 2001. pp 125-134.
- [8] Twidale, M., Randall, D., Bentley, R. (1994) "Situated Evaluation for Cooperative Systems". CSCW '94, 441-452.