

## *An Overview of Evaluation Methods for Collaborative Systems*

### *Um Panorama sobre Métodos de Avaliação de Sistemas Colaborativos*

Natália Sales Santos, Lidia Silva Ferreira e Raquel Oliveira Prates

Departamento de Ciência da Computação (DCC), Universidade Federal de Minas Gerais (UFMG).

Av. Antônio Carlos, 6627 – Pampulha, Belo Horizonte/MG, Brasil.

{nataliasales, lidiaferreira, rprates}@dcc.ufmg.br

**Abstract**—What is the state-of-the-art of the research on evaluation methods available for collaborative systems? Based on a Systematic Literature Review, this paper takes an initial step to answer that question and presents an overview of what has been published in the subject in the last 12 years.

**Resumo**—Qual o panorama atual de pesquisa em métodos de avaliação de sistemas colaborativos? A partir de uma revisão sistemática da literatura, o presente trabalho busca responder essa pergunta, apresentando uma visão geral da atual situação das avaliações dos sistemas colaborativos. Os resultados alcançados descrevem quais métodos têm sido utilizados nas avaliações, bem como os desafios de aplicá-los nesses contextos.

**Keywords**- evaluation methods, assessment

#### I. INTRODUÇÃO

A interface de um sistema é a parte com a qual uma pessoa entra em contato de forma física, perceptiva ou ainda conceitual, funcionando como o meio de interação com a aplicação [23]. Em outras palavras, é através da interface que o usuário entende e utiliza um sistema. O sucesso ou fracasso de um software está diretamente relacionado ao design de sua interface [33], tornando de fundamental importância avaliar a qualidade deste aspecto.

Existem muitos métodos de avaliação da qualidade de interfaces de sistemas monousuário, entretanto, não se pode dizer o mesmo para sistemas colaborativos. Isso porque, no contexto de sistemas colaborativos a avaliação é extremamente complicada, uma vez que, além de se considerar o comportamento e personalidade dos membros do grupo, o tempo gasto nesse tipo de avaliação pode ser altíssimo, já que as interações do grupo podem perdurar por dias ou até mesmo semanas. Logo, é necessário se identificar, quais aspectos devem ser avaliados, assim como os métodos a serem utilizados para a sua avaliação [3].

Alguns estudos foram realizados a fim de entender como são feitas as avaliações em sistemas colaborativos. Mesmo assim não existe consenso sobre o que avaliar, que método utilizar e quando avaliar [32]. Com isso existe a necessidade de continuar a investigação sobre os métodos de avaliação mais adequados a diferentes sistemas colaborativos e assim obter conhecimentos sobre os problemas encontrados durante as avaliações. Isso irá auxiliar os projetistas nas suas decisões sobre que métodos utilizar para avaliar os sistemas desenvolvidos.

Dada à necessidade de se conhecer o panorama das avaliações de sistemas colaborativos, este trabalho tem como

objetivo apresentar uma revisão sistemática da literatura que buscou investigar como esse tipo de sistema tem sido avaliado nos últimos anos.

Os resultados alcançados descrevem quais métodos têm sido utilizados nessas avaliações, bem como a forma que tem se tratado o desafio de avaliar sistemas colaborativos nos últimos anos. Na próxima seção, apresentamos os trabalhos já realizados, a fim de saber o que já se tem sobre avaliação de sistemas colaborativos. Em seguida, na seção Revisão Sistemática da Literatura, é apresentada a metodologia de pesquisa desse estudo. Na seção Artigos Selecionados são apresentados os resultados obtidos na pesquisa, seguida pela seção em que é feita a análise dos resultados obtidos. Logo após, na seção Discussão, respondemos à questão de pesquisa deste trabalho. Por fim, a seção Conclusão apresenta as considerações finais do estudo.

#### II. TRABALHOS RELACIONADOS

Foram encontrados poucos trabalhos que façam uma revisão da literatura de avaliação de sistemas colaborativos. O primeiro trabalho encontrado foi o artigo escrito por Pinelle e Gutwin [27], onde é apresentada uma revisão de trabalhos que avaliaram sistemas colaborativos da conferência ACM CSCW de 1990-1998. Nesse estudo foram analisados todos os artigos que introduziram ou avaliaram aplicações de groupware presentes na conferência, num total de quarenta e cinco artigos. Os autores propuseram uma forma de classificar os trabalhos dividindo-os em cinco áreas: tipo de avaliação, características da avaliação, coleta de dados e técnicas, avaliação do ciclo de desenvolvimento de software e tipo de conclusões tiradas a partir da avaliação.

Outro trabalho é de Wainer e Barsottini [38] que revisa artigos publicados também na conferência ACM CSCW, de 1998 a 2004. Este trabalho começou como uma tentativa de repetir a influência da “Review of groupware evaluation” por Pinelle e Gutwin [27]. No entanto, o escopo deste trabalho não se ateve a métodos de avaliação. Ao longo do trabalho perceberam que a avaliação de groupware é apenas uma das formas de pesquisa empírica e focaram neste tipo de pesquisa na área de CSCW.

Embora os trabalhos mencionados apresentem revisões da literatura sobre avaliação de sistemas colaborativos, eles analisaram apenas os artigos da ACM CSCW, sendo que o último trabalho publicado sobre o tema foi apresentado a oito anos. Com o uso cada vez maior dos sistemas colaborativos pela sociedade em diversos domínios, torna-se cada vez mais importante avaliar a qualidade dos sistemas sendo

desenvolvidos. Motivados pela falta de trabalhos recentes na área e pelo fato de que a avaliação de sistemas colaborativos continua sendo uma questão de pesquisa aberta (Grudin & Poltrack, [11]), a revisão proposta em nosso trabalho é relevante, pois descreve um panorama atualizado (até 2012) das pesquisas relacionadas à avaliação de sistemas colaborativos. Neste trabalho incluímos repositórios internacionais e nacionais, pois consideramos relevantes os trabalhos sobre o tema realizados em âmbito nacional.

### III. REVISÃO SISTEMÁTICA DA LITERATURA

Um método de revisão sistemática da literatura para pesquisas em Engenharia de Software denominado *Systematic Literature Review* (SLR) é apresentado por Kitchenham e outros [18][19]. Essa SLR é uma forma de identificar, avaliar e interpretar um conjunto amplo das pesquisas disponíveis relevantes para uma questão de pesquisa específica, área temática, ou fenômeno de interesse.

O uso de uma SLR permite ao pesquisador estruturar e organizar as informações obtidas pelo estudo. Além disso, possibilita a descrição dos trabalhos utilizados para pesquisa, bem como as bases de dados consideradas. Para realizar uma SLR três etapas devem ser seguidas:

- 1) Planejamento: etapa onde se faz a identificação das questões de pesquisa e a elaboração do protocolo de revisão;
- 2) Condução: etapa onde ocorre a seleção dos estudos, que é guiada pelo protocolo elaborado na primeira etapa;
- 3) Relatório: etapa onde acontece a validação e análise dos resultados.

Para averiguar os aspectos que devem ser considerados na seleção de um método de avaliação para sistemas colaborativos, foi utilizado como base o relatório técnico de Kitchenham [18][19] que apresenta procedimentos para a realização de uma SLR na Engenharia de Software.

As subseções abaixo apresentam as questões de pesquisa e o processo de pesquisa adotado.

#### A. Questão de pesquisa

Esta pesquisa tem como objetivo realizar um levantamento de trabalhos que realizam avaliações em sistemas colaborativos, a fim de obter o cenário geral das avaliações desses sistemas. Esse levantamento ajudou a responder a principal questão de pesquisa [QP] desse trabalho: *[QP1] Qual o panorama atual de pesquisa em métodos de avaliação de sistemas colaborativos?*

Para ajudar a responder essa questão foram criadas algumas perguntas específicas:

[QE1] *Quais métodos de avaliação estão sendo utilizados para realizar avaliações em sistemas colaborativos?*

[QE2] *Qual o custo/benefício de se aplicar o método de avaliação escolhido para avaliar um sistema colaborativo?*

[QE3] *Quais as dificuldades para se avaliar sistemas colaborativos?*

#### B. Processo de pesquisa

A nossa pesquisa fez o levantamento dos trabalhos publicados entre 2000 e 2012, ou seja, nos últimos 12 anos. A seleção deste intervalo foi baseada no ano de publicação da revisão feita por Pinelle e Gutwin [27].

A escolha das bases de dados incluiu os principais repositórios digitais da área de computação: IEEE, ACM, Science Direct e Springer; o repositório da área de Interação Humano-Computador (IHC), HCIBIB; os eventos da SBC das áreas de IHC e Sistemas Colaborativos: IHC e SBSC; o evento Collaboration Research International Workshop on Groupware (CRIWG) e os trabalhos do Núcleo de Pesquisa em Engenharia Semiótica e Interação (PENSI<sup>1</sup>), que é o grupo de pesquisa do Departamento de Ciência da Computação da Universidade Federal de Minas Gerais na área de IHC (ver TABELA I.).

Apesar de sabermos que outras bases de dados podem conter publicações sobre avaliações em sistemas colaborativos, consideramos que as bases selecionadas representam os principais repositórios de eventos e periódicos internacionais e de eventos nacionais de interesse.

Para a pesquisa automática nas bases de dados selecionadas foi utilizada a seguinte string de pesquisa: "groupware evaluation" or "collaborative systems evaluation" or "cooperative systems evaluation" or "CSCW evaluation" or "groupware assessment" or "collaborative systems assessment" or "cooperative systems assessment" or "CSCW assessment".

Para os Simpósios do IHC e SBSC cujos artigos não estavam disponíveis nos repositórios digitais pesquisados à época, as buscas foram manuais, através dos anais dos eventos. As pesquisas se deram da seguinte forma:

- Simpósio de Fatores Humanos em Sistemas Computacionais (IHC): todos os trabalhos publicados nesta conferência nos anos 2000 a 2004 e 2011.
- Simpósio Brasileiro de Sistemas Colaborativos (SBSC): todos os trabalhos publicados nesta conferência em 2004 e nos anos de 2006 e 2007.

TABELA I. BASE DE DADOS UTILIZADA PARA OBTER OS TRABALHOS PARA A REALIZAÇÃO DA SLR

Bibliotecas	Forma de pesquisa
IEEE, ACM, Science Direct e Springer	Busca Automática de 2000 a 2012.
HCIBIB	Busca Automática de 2000 a 2012.
IHC <sup>2</sup>	Busca Manual de 2000 a 2004 e 2011. Busca Automática a partir de 2006, 2008 e 2010.
SBSC	Busca Manual em 2004, 2006 e 2007. Busca Automática em 2005 e a partir de 2008.
CRIWG	Busca Automática de 2000 a 2011.

<sup>1</sup> [www.pensi.dcc.ufmg.br](http://www.pensi.dcc.ufmg.br)

<sup>2</sup> Vale notar que o IHC foi bianual de 2002 a 2010 voltando a ser anual a partir de 2011.

A metodologia utilizada para realizar a pesquisa é composta por quatro fases:

- Fase 1: eliminação por título;
- Fase 2: eliminação por resumo;
- Fase 3: eliminação por leitura diagonal;
- Fase 4: eliminação por leitura completa.

A eliminação em cada uma das fases foi feita através da aplicação de critérios de exclusão, a fim de manter o maior número de trabalhos relevantes. Os **critérios de exclusão** foram:

- Artigos de outras línguas diferentes de inglês, português e espanhol;
- Tutoriais, editoriais, pôsteres, painéis, palestras, mesas redondas, oficinas, demonstrações, workshops;
- Trabalhos duplicados que falam sobre o mesmo estudo serão considerados como equivalentes, sendo que o artigo mais recente é o que será utilizado como base para a análise.
- Trabalhos em que o estudo não avalia aspectos de interface/usabilidade de um sistema colaborativo.

A TABELA II. ilustra a quantidade de artigos inicial e que foram mantidos em cada uma das fases.

TABELA II. ARTIGOS SELECIONADOS NAS ETAPAS DA SLR

	Pesquisa	Etapa 1	Etapa 2	Etapa 3	Etapa 4
<b>ACM</b>	28	16	15	12	12
<b>IEEE</b>	14	12	12	4	3
<b>IHC</b>	12	12	10	1	1
<b>SBSC</b>	15	9	2	1	1
<b>Science Direct</b>	8	2	1	0	0
<b>Springer</b>	11	6	5	4	3
<b>HCIBIB</b>	44	24	23	9	7
<b>CRIWG</b>	8	7	6	4	1
<b>Grupo PENSi</b>	1	1	1	1	1
<b>TOTAL</b>	141	89	75	36	29

Na Fase 4, à medida que a leitura completa dos artigos era realizada, dados eram coletados para fins de análise. Para isso, foi criada uma lista de itens que deveriam ser coletados. O objetivo dessa lista era condensar as informações relevantes de forma a garantir a padronização da coleta [18][19]. As informações coletadas foram baseadas nas questões de pesquisa e estão apresentadas na TABELA III.

TABELA III. ITENS DA COLETA DE DADOS

<b>Dados gerais</b>	Titulo; fonte; autores; local e ano de publicação.
<b>Dados da Pesquisa</b>	Objetivo; contexto; aspecto avaliado, motivação; critério para escolha do sistema e do método de avaliação; dificuldades, vantagens e desvantagens encontradas na aplicação do método; quantos avaliadores e usuários foram utilizados; resultados; informações consideradas para análise, tipo do método (novo, adaptado, existente); tipo de coleta; momento; tipo de análise; domínio da avaliação; tipo de tecnologia foi utilizada (geral ou específica); ambiente do teste (real ou controlado); foco do estudo (método ou sistema); passos do método para avaliação.

Na Fase 4, que envolve a leitura completa do artigo, definem-se critérios de qualidade que permitem a definição de se o artigo atende ao objetivo da pesquisa ou não. Cada critério foi avaliado como “sim”, “parcial” ou “não” e foram pontuados com o valor de 1, 0,5 e 0, respectivamente. Cada critério também possui um peso, que corresponde ao grau de importância (1 – baixo, 2 – médio, ou 3 – alto) deste critério para responder as questões de pesquisa (ver TABELA IV. ). Artigos que obtiveram uma nota menor que 50% nos critérios de qualidade foram eliminados nesta última fase.

TABELA IV. CRITÉRIOS DE QUALIDADE

Critérios de Qualidade	Peso
O estudo define claramente o objetivo da pesquisa (define questão de pesquisa)?	1
O artigo responde às questões de pesquisa definidas?	1
O artigo relata os passos do método que foram seguidos para a avaliação?	2
O artigo cita o contexto do sistema colaborativo em análise?	3
O artigo apresenta a motivação para a escolha do método?	1
Apresenta dificuldades para avaliar o sistema colaborativo?	2
O artigo apresenta o custo/benefício de utilizar o método estudado?	3
O artigo mostra quantos usuários foram necessários para realizar o estudo?	2
<b>TOTAL</b>	<b>15</b>

Algumas precauções foram tomadas a fim de evitar erros na execução da SLR. Além de realizar a revisão da literatura com dois avaliadores, tomou-se por definição que em caso de dúvida se um artigo deveria ou não ser eliminado, o mesmo deveria ser mantido para análise na etapa seguinte.

#### IV. ARTIGOS SELECIONADOS

Durante o estudo foram selecionados 29 trabalhos relacionados à avaliação de sistemas colaborativos. Os métodos apresentados pelos trabalhos selecionados foram classificados e são apresentados na TABELA V. , bem como as suas devidas referências. Os métodos identificados foram classificados de acordo com cinco dimensões: origem, tipo de coleta, momento da coleta, foco e tipo de análise. A **origem** do método se refere de onde ele vem, ou seja, se é um método novo, se é um método existente, ou se é uma adaptação de algum método existente (na próxima seção fazemos uma discussão mais detalhada desta dimensão). Para o **tipo de coleta** analisamos se a avaliação foi feita através da coleta de dados direta do usuário (e.g questionários e entrevistas), inspeção (quando um especialista inspecciona o sistema identificando os problemas), ambiente controlado (testes com usuários são realizados em um ambiente controlado, como por exemplo, um laboratório de teste) ou contexto real (a coleta de dados é feita no contexto real do usuário). O **momento da coleta** classifica a avaliação feita como formativa – i.e. durante o processo de desenvolvimento; ou somativa, avaliação feita com o software pronto.

TABELA V. REFERÊNCIAS DOS TRABALHOS ANALISADOS

ID	Método	Origem	Tipo de Coleta	Momento	Foco	Tipo de análise
[36]	Avaliação envolvendo usuários	EX:MU	Ambiente controlado	Avaliação somativa	Método	Qualitativa
[32]	Avaliação envolvendo usuários	EX:MU	Ambiente controlado	Avaliação formativa	Método	Qualitativa
[7]	Avaliação envolvendo usuários	A:CE-CG	Ambiente real	Avaliação formativa e somativa	Sistema	Qualitativa
[39]	Avaliação envolvendo usuários	EX:MU	Ambiente controlado	Avaliação somativa	Sistema	Quantitativa
[4]	Avaliação Heuristica para Groupware	A:MU-CG	Inspeção	Avaliação somativa	Método	Qualitativa
[28]	Cognitive walkthrough	A:MU-CG	Inspeção	Avaliação formativa	Método	Qualitativa
[41]	Coleta de dados dos usuários e questionários	EX:MU	Ambiente real	Avaliação somativa	Sistema	Qualitativa
[35]	Coleta de dados dos usuários e questionários	EX:MU	Ambiente real	Avaliação somativa	Sistema	Qualitativa
[2]	Coleta de dados dos usuários e questionários	EX:MU	Questionários	Avaliação somativa	Sistema	Qualitativa
[8]	Combinação de métodos	A:MU-CE	Inspeção	Avaliação formativa e somativa	Método	Qualitativa
[1]	Cooperation Scenarios (COS)	A:CG-CE	Ambiente controlado	Avaliação somativa	Método	Qualitativa
[29]	CUA (Collaboration Usability Analysis)	NOVO	Ambiente real	Avaliação formativa	Método	Qualitativa
[17]	E-MAGINE	NOVO	Ambiente controlado	Avaliação formativa e somativa	Método	Qualitativa
[25]	Framework de avaliação multifacetada	NOVO	Ambiente real	Avaliação somativa	Método	Qualitativa
[6]	Framework for Evaluating Groupware Systems	A:CG-CE	Inspeção	Ambiente controlado	Método	Qualitativa
[36]	Heurísticas de Baker	EX:C	Inspeção	Avaliação somativa	Método	Qualitativa
[32]	Inspeção informal	A:MU-CG	Inspeção	Avaliação formativa	Método	Qualitativa
[1]	Keystroke-Level Model (KLM)	A:MU-CG	Coleta de dados do usuário (entrevista e questionário)	Avaliação somativa	Método	Qualitativa
[16]	Lifecycle-based e entrevistas	A:MU-CG	Ambiente controlado	Avaliação formativa e somativa	Sistema	Qualitativa
[5]	Método baseado no SPADA	A:CE-CG	Ambiente real	Avaliação somativa	Método	Qualitativa
[22]	Método de Avaliação de Comunicabilidade para sistemas colaborativos (MACg)	A:MU-CG	Ambiente controlado	Avaliação somativa	Método	Qualitativa
[34]	Método de Inspeção Semiótica (MIS)	EX:MU	Inspeção	Avaliação somativa	Sistema	Qualitativa
[22]	Método de Inspeção Semiótica (MIS)	EX:MU	Inspeção	Avaliação formativa e somativa	Método	Qualitativa
[37]	Método de interação colaborativa em um ambiente móvel compartilhada, em condições visuais e os olhos livres	NOVO	Ambiente controlado	Avaliação somativa	Sistema	Qualitativa
[31]	Networked Game Heuristics (NGH)	A:MU-CE	Inspeção	Avaliação somativa	Método	Qualitativa
[40]	Observação em ambiente controlado	EX:MU	Ambiente controlado	Avaliação formativa e somativa	Sistema	Qualitativa
[12]	Observação em ambiente real	EX:MU	Ambiente real	Avaliação formativa e somativa	Método	Qualitativa
[15]	Observação em ambiente real e questionários	EX:MU	Ambiente real	Avaliação formativa e somativa	Sistema	Qualitativa
[13]	Scenario-based evaluation (SBE)	NOVO	Coleta de dados do usuário (entrevista e questionário)	Avaliação somativa	Método	Qualitativa
[24]	SMAM	NOVO	Ambiente controlado	Avaliação somativa	Método	Quantitativa
[14]	SWIMs: Scenario Walkthrough e Método de Inspeção	A:MU-CE	Inspeção	Avaliação formativa e somativa	Método	Qualitativa
[30]	Técnica de inspeção de usabilidade T-CUA	A:CG-CE	Inspeção	Avaliação formativa e somativa	Método	Qualitativa

O **foco** classifica os estudos em dois tipos: os estudos que tem como objetivo apresentar um método e os que têm como objetivo apresentar um sistema. Os que têm foco no método realizam uma avaliação para testar/mostrar como o método funciona; já os que têm foco no sistema, realizam uma avaliação do sistema para saber o quanto bom o sistema é para o propósito no qual ele foi desenvolvido. Finalmente, o tipo de análise classifica a análise dos dados feita como quantitativa ou qualitativa. Assim, na TABELA V. , listamos os métodos estudados em cada trabalho e sua classificação.

## V. ANÁLISE

Para respondermos nossa questão principal de pesquisa (QP1), os dados coletados foram analisados a partir das respostas das questões específicas (QE). Para a análise, coletamos de cada artigo em estudo os dados listados na TABELA IV. , que foram coletados durante a Fase 4 da SLR. A seguir apresentamos cada uma das três questões específicas e a resposta obtida através da análise feita. Com base nestas respostas, passamos então para a análise da questão mais geral de pesquisa (QP1) sobre o panorama geral de avaliação de sistemas colaborativos apresentada na próxima seção.

*[QE1] Quais métodos de avaliação estão sendo utilizados para realizar avaliações em sistemas colaborativos?*

Para responder a essa pergunta analisamos os métodos em relação à sua origem. Conforme explicado na seção anterior, nesta dimensão os métodos foram classificados em três categorias:

- Novo: quando o objetivo do artigo é apresentar o método e não há referências a publicações anteriores sobre ele.
- Existente: quando os passos para a aplicação de um método já existente (publicado anteriormente) não são alterados.
- Adaptado: quando é utilizado um método existente com modificações em sua aplicação.

Essa classificação se fez necessária para analisarmos quais métodos já consolidados estão sendo utilizados e em quais contextos surge a necessidade de métodos novos ou de adaptações. Foi observado que a metade (50%) dos trabalhos utilizou métodos adaptados. A apresentação de métodos novos é feita em 16% dos artigos, enquanto o restante dos artigos (34%) apresentam métodos existentes. A maior utilização de métodos adaptados se deve ao fato de que muitas vezes o método escolhido para avaliação não é voltado para sistemas colaborativos, logo, as adaptações se fazem necessárias para que o método possa contemplar a parte colaborativa do sistema avaliado. Para ilustrar, tomemos o trabalho de Baker, Greenberg e Gutwin [4]. Neste trabalho, os autores apresentam um conjunto de heurísticas específicas para sistemas colaborativos. Eles propõem o uso do método da Avaliação Heurística proposta por Nielsen [26], porém usando este novo conjunto de heurísticas em substituição às heurísticas originais de Nielsen.

Nos trabalhos que apresentam métodos novos, em sua maioria, os autores além de apresentar o método, também

descrevem um estudo de caso que tem por objetivo demonstrar que esse novo método é capaz de avaliar sistemas colaborativos e gerar informações relevantes para tal. Ao longo do estudo percebemos que alguns desses métodos novos foram desenvolvidos exclusivamente para avaliar um determinado sistema. Trabalhos como o de Pinelle e Gutwin [30] mostra que o método criado para avaliar uma tecnologia específica se mostrou mais eficiente que qualquer outro método de inspeção, porém, não se sabe se os mesmos resultados serão encontrados caso o método seja utilizado para avaliar um sistema colaborativo em outro contexto.

Os artigos sobre métodos existentes em sua maioria foram utilizados para verificar a aplicabilidade do método, ou seja, se o mesmo pode ser aplicado a sistemas colaborativos sem haver a necessidade de alterações. A motivação para estes estudos seria identificar o benefício no uso do método, uma vez que não necessitaria uma proposta de adaptação ou o seu aprendizado pelos avaliadores.

Na tentativa de se analisar o quanto consolidado um método novo ou adaptado se tornou, percebe-se que normalmente eles são tratados apenas na publicação que os apresenta, ou no máximo em uma que os apresenta e outra que os avalia (muitas vezes em comparação com outros) e normalmente de autoria dos mesmos pesquisadores ou grupo que propuseram o método. Isto é um problema, pois o método é apresentado, mas não se tem uma avaliação ampla deste método indicando seus benefícios e limitações em diferentes contextos de uso, o que dificulta a consolidação de novos métodos. No entanto, este problema pode ser causado pela própria comunidade científica que tipicamente não julga que trabalhos de “replicação”, que seriam necessários para aplicar e confirmar estes métodos sejam originais e não os aceitam para publicação [9].

Um aspecto importante a ser analisado é o domínio em que os métodos de avaliação apresentados são aplicados. Para a presente pesquisa foi criada uma definição de domínio, classificado em dois tipos:

- Domínio geral: métodos que se aplicam a sistemas colaborativos cujo domínio de aplicação é geral, ou seja, é definido no protocolo social.
- Domínio específico: métodos que levam em consideração características específicas de um domínio para avaliar sistemas colaborativos cujo domínio de aplicação é definido no protocolo tecnológico.

Através da Fig. 1, observamos a distribuição dos artigos em relação à origem dos métodos ao longo do tempo. Inicialmente classificamos os métodos nos três tipos principais: novo, adaptado e existente. Mas para melhor entender os tipos de adaptações realizadas e a origem dos métodos existentes, fazemos uma análise mais detalhada dos métodos adaptados e existentes. Para os adaptados identificamos se o método original era para avaliar sistemas monousuários ou já voltado para sistemas colaborativos (DE) e se a adaptação teve o objetivo de adaptá-lo para sistemas colaborativos ou para um domínio específico (PARA) (e.g. sistemas colaborativos educacionais). Abaixo descrevemos as subcategorias:

- A:MU-CG: ADAPTADO: DE monousuário PARA colaborativo genérico: onde o método de avaliação foi desenvolvido originalmente para a avaliação de sistemas monousuários e foi adaptado para realizar a avaliação de sistemas colaborativos genérico.
- A:C-CG: ADAPTADO: DE colaborativo específico PARA colaborativo genérico: onde o método de avaliação foi desenvolvido originalmente para a avaliação de sistemas colaborativos e foi adaptado para realizar a avaliação de sistemas colaborativos genérico.
- A:MU-CE: ADAPTADO: DE monousuário PARA colaborativo específico: onde o método de avaliação foi desenvolvido originalmente para a avaliação de sistemas monousuários e foi adaptado para realizar a avaliação de sistemas colaborativos de domínio específico.
- A:C-CE: ADAPTADO: DE colaborativo genérico PARA colaborativo específico: onde o método de avaliação foi desenvolvido originalmente para a avaliação de sistemas colaborativos e foi adaptado para realizar a avaliação de sistemas colaborativos específicos.
- EX:MU: EXISTENTE: DE monousuário: onde o método de avaliação foi desenvolvido originalmente para a avaliação de sistemas monousuários.
- EX:C: EXISTENTE: DE colaborativo: onde o método de avaliação foi desenvolvido originalmente para a avaliação de sistemas colaborativos.

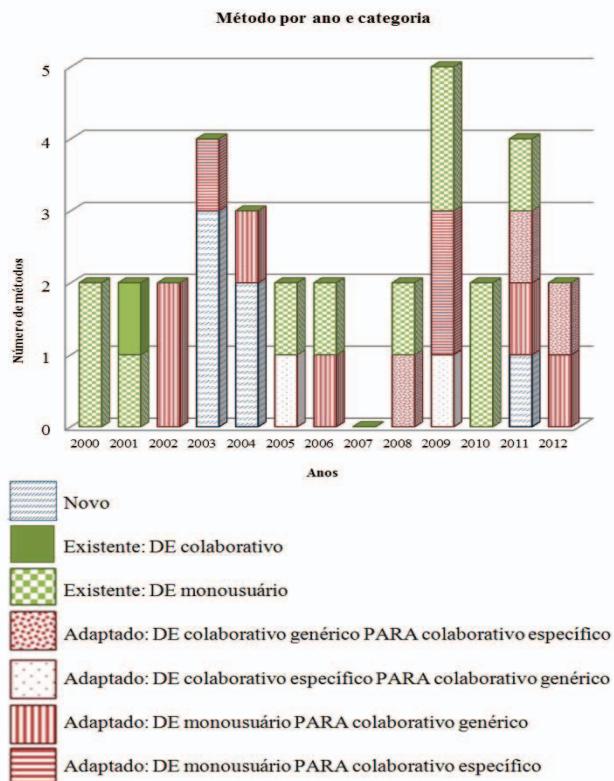


Figura 1. Métodos de avaliação distribuídos por anos e categorias

Observando o panorama dos últimos doze anos (Fig. 1), percebemos que ao longo dos anos alguns métodos novos foram propostos. O maior número de pesquisas desses métodos foi em 2003 e 2004, ocorrendo também em 2011. Já sobre os métodos adaptados, percebemos que a adaptação mais recorrente é a que se refere a métodos originalmente criados para avaliar sistemas monousuários que foram adaptados para aplicações colaborativas genéricas.

Observamos que, no período analisado, adaptações de métodos colaborativos genéricos para domínios específicos são bem menores que os trabalhos de adaptação de métodos monousuários para o contexto colaborativo genérico. Por fim, percebemos através do gráfico, que métodos de avaliação adaptados provindos de sistemas monousuários foram os mais utilizados em todo o período de tempo abrangido por essa pesquisa. Isso se deve ao fato de existirem poucos métodos para a avaliação de sistemas colaborativos, o que leva os pesquisadores a adaptar métodos originalmente criados para a avaliação de sistemas monousuários para realizar suas avaliações.

Em relação aos métodos existentes para sistemas monousuários percebe-se que, apesar da proposta de métodos específicos para sistemas colaborativos, eles continuam sendo usados ao longo do período analisado. O uso dos métodos colaborativos existentes não apresenta um crescimento expressivo em relação aos outros métodos. A única utilização deste método ocorre em 2001.

A Fig. 2, nos mostra que a maior parte das pesquisas sobre métodos (sejam eles novos, adaptados ou existentes) focou em sistemas de domínio geral. As pesquisas que realizam estudos sobre os métodos novos e adaptados, embora tenham sido mais focadas nos sistemas de domínio geral, tiveram uma maior ocorrência que os métodos existentes para o domínio específico. Tal situação é bastante plausível, uma vez que para se levar em consideração especificidades de um determinado domínio foi necessário adaptar o método para que a avaliação fosse efetiva ou precisou-se criar um método novo.

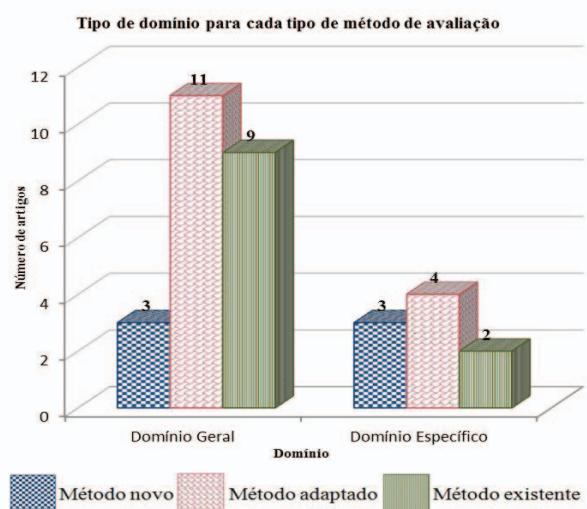


Figura 2. Domínio de aplicação dos métodos de avaliação

Analisamos também em relação às avaliações que envolveram testes com usuários, o ambiente em que esses métodos foram aplicados (ambiente controlado x ambiente real). Nos artigos analisados 75% dos métodos que envolveram testes com usuários foram conduzidos em ambientes controlados, enquanto que 25% conduziram testes em ambientes reais.

*[QE2] Qual o custo/benefício de se aplicar o método de avaliação escolhido para avaliar um sistema colaborativo?*

Nem todos os trabalhos citam o custo/benefício de se utilizar um determinado método, então para responder essa pergunta, apresentamos os dados colhidos dos trabalhos que apresentaram essa informação.

Os estudos que utilizaram métodos de inspeção têm como maior benefício, em sua maioria, o baixo custo de sua aplicação, pois não é necessária a utilização de usuários. E mesmo sendo o método de baixo custo, ele é eficiente, pois foi possível encontrar uma quantidade significativa de problemas nos sistemas avaliados. Este custo/benefício é o mesmo relatado para métodos de inspeção para sistemas monousuários, e podemos argumentar que já seria esperado. Poucos artigos mencionam outros custos específicos para sistemas colaborativos. No entanto, Prates e Raposo [32] indicam como custos adicionais deste tipo de método a necessidade de se avaliar o sistema para diferentes papéis do grupo e, também de se simular a interação entre os papéis para se avaliar os impactos das ações de um usuário em outro. Além disso, o avaliador pode ter dificuldades em avaliar partes síncronas do sistema, precisando, muitas vezes, de mais de um avaliador ou alguém (mesmo que não seja especialista), que participe da avaliação.

Na maioria dos estudos que utilizaram usuários para realizar a avaliação, obteve-se um alto custo de aplicação, pois conseguir usuários com disponibilidade de participar da avaliação pode ser um desafio.

Os trabalhos que utilizaram métodos novos em sua maioria não informaram o custo/benefício que tiverem em relação à aplicação do método, mas em alguns casos, como no método E-MAGINE apresentado em [17], o custo/benefício está relacionado à sua estrutura modular. Ele explicitamente visa determinar a fonte de sucesso ou fracasso de uma fase inicial do desenvolvimento do sistema. Outro estudo cita que o maior custo de um método novo é em relação ao tempo gasto para o aprendizado do método e a falta de consolidação dos resultados.

*[QE3] Quais os desafios para avaliar sistemas colaborativos?*

Assim como acontece com o custo/benefício da aplicação dos métodos, muitos estudos não citam se tiveram alguma dificuldade durante a sua aplicação. Mas os estudos que citaram, relataram que as dificuldades mais encontradas estão relacionadas à simulação do ambiente de colaboração, à quantidade de participantes (quando o método precisa de usuários) e ao alto custo de aplicação do método.

Em alguns estudos percebemos que algumas dessas dificuldades se devem ao fato de o método não contemplar a parte colaborativa do sistema, então é necessário realizar

algumas mudanças, porém nem sempre essas mudanças ajudam/melhoram a aplicação. Essas mudanças não configuram uma proposta de um método adaptado, mas são alterações feitas para o contexto específico para o estudo de caso conduzido.

## VI. DISCUSSÃO

A partir dos resultados obtidos é possível responder a questão de pesquisa investigada nesse trabalho: *[QPI] Qual o panorama atual de pesquisa em métodos de avaliação de sistemas colaborativos?*

No início dos anos 2000 o número de trabalhos na área de pesquisa em avaliação de sistemas colaborativos é baixo, porém, nos anos 2003, 2004, 2009 e 2011 nota-se um crescimento do número de pesquisas. O crescimento destas pesquisas pode estar relacionado ao fato de que os sistemas colaborativos têm sido cada vez mais incorporados ao cotidiano das pessoas. Podemos entender que esse aumento nas pesquisas é um indicador de que a questão sobre métodos que permitam avaliar sistemas colaborativos ainda é uma questão aberta de pesquisa.

Os estudos que têm como foco apresentar o método (ver TABELA V. ), mostram a necessidade e o investimento em se identificar métodos apropriados para sistemas colaborativos, sejam eles de domínio geral ou específico. Já os estudos que focam no sistema, apresentam a avaliação como forma de discutir a qualidade do sistema. Embora estes trabalhos normalmente não detalhem as informações sobre a aplicação do método, eles colaboraram para a consolidação dos métodos utilizados.

Como se sabe, não há um método que seja sempre melhor que os demais. Assim, na decisão de que método usar os avaliadores devem analisar os custos e benefícios de cada método. Os artigos investigados permitiram uma análise do custo e benefício relacionado ao envolvimento ou não do usuário no método. Embora esta discussão seja um primeiro passo na seleção do método a ser usado, é importante que se invista em pesquisas que façam análises comparativas de métodos e que informações sobre custo e benefício sejam incluídas nos artigos que discutem o uso de métodos.

## VII. CONCLUSÃO

Neste trabalho apresentamos um panorama de como sistemas colaborativos estão sendo avaliados. As informações foram obtidas através de uma revisão sistemática da literatura que possibilitou o levantamento dos trabalhos mais relevantes sobre o assunto. A motivação para essa análise partiu da falta de consolidação sobre quais métodos utilizar e em quais situações, e pelo fato de que trabalhos que fizeram uma discussão geral da pesquisa na área já foram publicados há mais de oito anos.

Com os resultados, foi possível mostrar que existe uma quantidade considerável de métodos que podem ser utilizados para avaliar sistemas colaborativos. Esses métodos podem ser classificados como novos, adaptados ou existentes. Os mais utilizados são os métodos adaptados, pois em grande parte esses métodos já são consolidados e permitem que, com algumas adaptações, passem a contemplar a parte colaborativa dos sistemas.

Há também uma quantidade significativa de uso de métodos novos, tanto para avaliar sistemas colaborativos em domínios específicos quanto gerais. Apesar de ocorrer menos métodos novos do que adaptados e existentes, a ocorrência desses métodos é notável, dado a complexidade envolvida em criar uma nova metodologia de avaliação para sistemas colaborativos. Metade dos métodos novos apresentados são para domínio geral de avaliação, o que colabora para ampliar o conjunto de métodos disponíveis para avaliar sistemas colaborativos de forma geral. Um problema que pode ser percebido na apresentação desses novos métodos é a falta de trabalhos que colaborem para sua consolidação através da sua aplicação em outros tipos de sistemas ou a contextos diferentes.

Em termos de contribuições, este trabalho serve como base para que outros pesquisadores conheçam o panorama sobre os métodos de avaliação em sistemas colaborativos e possam tanto discutir como suas pesquisas se encaixam neste panorama, quanto identificar questões de interesse. Para pesquisadores ou desenvolvedores que pretendam avaliar sistemas colaborativos, o levantamento dos métodos e sua classificação de acordo com o tipo de coleta, momento e análise, assim como a discussão sobre custos, benefícios e desafios podem auxiliá-los na definição de qual método seria mais adequado para seu contexto específico.

Um trabalho futuro de interesse para esta pesquisa é o desenvolvimento de um sistema que permitisse uma visualização (por método) sobre a análise feita neste trabalho. Com isso, pode-se permitir que uma pessoa obtenha uma descrição mais detalhada sobre cada método. De posse dessa informação seria possível comparar métodos de interesse. Outra direção interessante é ampliar a análise feita discutindo outros aspectos sobre os métodos, como por exemplo, se a aplicação dos métodos foi feita em contexto acadêmico ou real, ou detalhando aspectos específicos tratados em custo e benefício.

#### AGRADECIMENTOS

Agradecemos ao apoio parcial a esta pesquisa pela CAPES, FAPEMIG e também ao INCT-Web (INWeb) (MCT/CNPq/ proc 57.3871/2008-6).

#### REFERÊNCIAS

- [1] Antunes, P., Herskovic, V., Ochoa, S. F., Pino, J. A. (2008) Structuring dimensions for collaborative systems evaluation. In ACM Comput. Surv. 44, 2, Article 8 (March 2008), 28 pages.
- [2] Becker, K.; Bacelo, A.P.T.; , "The evaluation of GRADD: a GDSS supporting asynchronous and distributed meetings," Groupware, 2000. CRIWG 2000. Proceedings. Sixth International Workshop on , vol. no., pp.19-26, 2000.
- [3] Baker, K.; Greenberg, S. & Gutwin, C. (2001). Heuristic evaluation of groupware based on the mechanics of collaboration. Lecture Notes in Computer Science, pp. 123—140.
- [4] Baker, K., Greenberg S., Gutwin C. (2002) Empirical development of a heuristic evaluation methodology for shared workspace groupware. In Proc. of CSCW '02.
- [5] Burkhardt, J. M., Détienne, F., Hébert, A. M., Perron, L., Safin, S., Leclercq, P. (2009) An approach to assess the quality of collaboration in technology-mediated design situations. In Conf. of ECCE '09.
- [6] Gallardo, J., Molina, A. I. , Bravo, C., Redondo M. Á., Collazos, C. A. Empirical and Heuristic-Based Evaluation of Collaborative Modeling Systems: An Evaluation Framework. In Group Decision and Negotiation, 2011, Volume 20, Number 5, Pages 535-562
- [7] Gauducheau, N.; Soulier, E.; Lewkowicz, M. (2005) Design and evaluation of activity model- based groupware: methodological issues. In Proc. of 14th IEEE WETICE'05.
- [8] Goebbel, G., Lalioti, V.i, Göbel M. (2003) Design and evaluation of team work in distributed collaborative virtual environments. In Proc. of VRST '03.
- [9] Greenberg, S. and Buxton, B. (2008) Usability evaluation considered harmful (some of the time). In Proc. of CHI '08, (2008).
- [10] Grudin, J. (1988). Why groupware applications fail: Problems in design and evaluation. Office: Technology and People, 4(3), 245-264.
- [11] Grudin, J., Poltrock, S. (2012) CSCW - Computer Supported Cooperative Work. In: Soegaard, Mads and Dam, Rikke Friis (eds.), "Encyclopedia of Human-Computer Interaction". Aarhus, Denmark: The Interaction-Design.org Foundation.
- [12] Guy E. S. (2005). "...real, concrete facts about what works..." integrating evaluation and design through patterns. In Proc. of GROUP '05.
- [13] Haynes, S. R., Purao, S., Skattebo, A. L. (2004) Situating evaluation in scenarios of use. In Proc. of CSCW '04.
- [14] Haynes, S. R., Purao, S., Skattebo, A. L., Haynes, S. R. (2009) Scenario-Based Methods for Evaluating Collaborative Systems. In CSCW, Volume 18, Number 4 (2009), 331-356.
- [15] Hirano, S. H., Yeganyan, M. T., Marcu, G., Nguyen, D. H., Boyd, L. A., Hayes, G. R. (2010) vSked: evaluation of a system to support classroom activities for children with autism. In Proc. of 28th CHI '10.
- [16] Huang, J.P.H. (2004) An enhanced approach to support collaborative systems evaluation. In Proc. of IEEE 13th WETICE 2004.
- [17] Huis in 't Veld, M.A.A.; Andriessen, J.H.E.; Verburg, R.M. (2003) E-MAGINE: the development of an evaluation method to assess groupware applications. In Proc. of 12th IEEE WETICE 2003.
- [18] Kitchenham, B. A. Procedures for Undertaking Systematic Reviews, Joint Technical Report, Computer Science Department, Keele University (TR/SE-0401) and National ICT Australia Ltd. ( 0400011T.1), 2004.
- [19] Kitchenham, B. A., Dybå, T., Jørgensen, M. Evidence-based software engineering. In: Proceedings of the 26th International Conference on Software Engineering, (ICSE'04), IEEE Computer Society, Washington DC, USA, 2004, pp. 273–281.
- [20] Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., Linkman, S. Systematic literature reviews in software engineering - A systematic literature review, Information and Software Technology, Volume 51, Issue 1, Special Section - Most Cited Articles in 2002 and Regular Research Papers, January 2009, Pages 7-15.
- [21] Mattos, B.A.M.; Prates, R.O.; An Overview of the Communicability Evaluation Method for Collaborative Systems. In: IAÁDIS International Conference WWW/Internet 2011, 2011, Rio de Janeiro. Proceedings of WWW/Internet 2011, 2011. p. 129-136.
- [22] Mattos, B.A.M.; Santos, R.L.; Prates, R.O.; , "Investigating the Applicability of the Semiotic Inspection Method to Collaborative Systems," Sistemas Colaborativos (SBSC),

- 2009 Simposio Brasileiro de , vol., no., pp.53-60, 5-7 Oct. 2009.
- [23] Moran, T. (1981). The Command Language Grammars: a representation for the user interface or interactive computer systems. In International Journal of Man-Machine Studies 15:3-50, Academic Press.
- [24] Mühlfelder, M., Luczak, H. (2003) Cognitive Analysis of Process Knowledge Transfer in Computer Supported Cooperative Work. In International Journal of Human-Computer Interaction.
- [25] Neale, D. C., Carroll, J. M., Rosson. M. B. (2004) Evaluating computer-supported cooperative work: models and frameworks. In Proc of CSCW '04.
- [26] Nielsen, J. (1993) Usability heuristics. In Nielsen, J. (Ed) Usability Engineering. Academic Press. 115-164.
- [27] Pinelle, D. Gutwin, C. (2000) A review of groupware evaluations. In Proc. of 9th IEEE WETICE 2000.
- [28] Pinelle, D., Gutwin C. (2002) Groupware walkthrough: adding context to groupware usability evaluation. In Proc. of CHI '02.
- [29] Pinelle, D., Gutwin, C., Greenberg, S. (2003) Task analysis for groupware usability evaluation: Modeling shared-workspace tasks with the mechanics of collaboration. In ACM Trans. Comput.-Hum. Interact. 10, 4 (December 2003), 281-311.
- [30] Pinelle, D., Gutwin C. (2008) Evaluating teamwork support in tabletop groupware applications using collaboration usability analysis. In Personal Ubiquitous Comput. 12, 3 (January 2008), 237-254.
- [31] Pinelle, D., Wong, N., Stach, T., Gutwin, C. (2009). Usability heuristics for networked multiplayer games. In Proc. of GROUP '09.
- [32] Prates, R. O., Raposo, A. B. (2006). Desafios para testes de usuários em sistemas colaborativos - lições de um estudo de caso. In Proc.of VII IHC '06.
- [33] Preece, J.; Rogers, Y.; Sharp, E. (2007) Interaction Design: Beyond Human Computer Interaction. 2nd ed. Hoboken, New Jersey: Wiley.
- [34] Santos, R. L., Prates, R. O. (2010) Estratégias para comunicar qualidade na Wikipedia. In Proc. of IX IHC '10.
- [35] Sohlenkamp, M., Prinz W., Fuchs, L. (2000) PoLI awaC: design and evaluation of an awareness-enhanced groupware client. In Proc of AI & SOCIETY 2000.
- [36] Steves, M. P., Morse E., Gutwin C., Greenberg S. (2001) A comparison of usage evaluation and inspection methods for assessing groupware usability. In Proc. of GROUP '01.
- [37] Tredafilov, D., Vazquez-Alvarez, Y., Lemmelä, S., Murray-Smith, R. (2011) "Can we work this out?": an evaluation of remote collaborative interaction in a mobile shared environment. In Proc. of 13th MobileHCI '11.
- [38] Wainer, J., Barsottini C. Empirical research in CSCW — a review of the ACM/CSCW conferences from 1998 to 2004. In Journal Of The Brazilian Computer Society. Volume 13, Number 3 (2007), 27-35.
- [39] Wallace, J. R., Scott, S. D., Stutz T., Enns T., Inkpen, K. (2009) Investigating teamwork and taskwork in single- and multi-display groupware systems. In Personal Ubiquitous Comput. 13, 8 (November 2009), 569-581.
- [40] Yousef , K. (2008) Preliminary evaluation of a remote mobile collaborative environment. In CHI '08 extended abstracts on Human factors in computing systems (CHI EA '08).
- [41] Zhou, Y., Percival, G., Wang, X., Wang, Y., Zhao, S. (2011) MOGCLASS: evaluation of a collaborative system of mobile devices for classroom music education of young children. In Proc. of CHI '11.