Online em: www2.serg.inf.puc-rio.br/docs/PratesLeitaoFigueiredo-IHC2004.pdf

Desafios de Avaliação de Interfaces de Ambientes Educacionais – Um Estudo de Caso

Raquel Oliveira Prates
Departamento de Informática
IME – UERJ
R. São Francisco Xavier, 524 – 6o. andar, Maracanã, RJ
raquel@ime.uerj.br

Carla Faria Leitão
Departamento de Psicologia,
SERG/DI - PUC-Rio
R. Marquês de São Vicente, 225
Gávea, RJ
cfaria@inf.puc-rio.br

Rosa Maria V. de Figueiredo
Departamento de Matemática Aplicada – IME – UERJ
R. São Francisco Xavier, 524 – 6o. andar, Maracanã, RJ
rosa@ime.uerj.br

RESUMO
Ambientes educacionais trazem novos desafios para o design e avaliação de interfaces [2, 10]. Neste artigo levantamos questões sobre a avaliação de interfaces educacionais, a partir de um estudo de caso. Este estudo compreendeu a avaliação formativa da interface de um sistema de apoio ao aprendizado utilizando o método de avaliação de comunicabilidade [4, 5].

Palavras Chave
Avaliação, comunicabilidade, interfaces educacionais.

INTRODUÇÃO
Algumas teorias, métodos e técnicas já estão bem consolidados para a área de IHC [7]. No entanto, alguns domínios têm especificidades que podem tornar necessária a adaptação ou mesmo a revisão destes métodos e técnicas e apresentam novos desafios para o design e avaliação de suas interfaces. Este parece ser o caso do domínio educacional, no qual, por exemplo, uma interface pode deliberadamente permitir (ou até mesmo induzir) o erro do usuário para propiciar ou consolidar seu aprendizado [2]. Neste contexto a máxima de IHC de que a interface deve sempre prevenir erros de seus usuários [3, 8] é conscientemente violada.

À medida que ambientes educacionais vêm sendo, cada vez mais, integrados às práticas de ensino e aprendizado justifica-se investir na pesquisa sobre como tratar as particularidades do domínio educacional. Tal investimento já começa a se fazer presente na área de IHC por meio, por exemplo, de propostas de métodos de avaliação preditiva capazes de auxiliar o professor na escolha de software adequados a seus objetivos didáticos [10].

Neste artigo, adotamos um foco distinto ao discutir a avaliação formativa da interface de ambientes educacionais com o objetivo de melhorar sua qualidade. Para isto, relatamos a nossa experiência com a avaliação formativa de um ambiente de apoio ao aprendizado utilizando o método de avaliação de comunicabilidade [4, 5], e levantamos algumas questões específicas encontradas nesta avaliação. Estas questões são relevantes, uma vez que ilustram os desafios de design de interfaces no contexto educacional e criam a demanda para uma investigação mais profunda sobre a necessidade de adaptação de métodos de avaliação existentes para este domínio ou da proposta de novos métodos específicos.

Na próxima seção apresentamos o sistema Ampliar [6] utilizado neste estudo de caso. A seguir apresentamos o método de avaliação de comunicabilidade e o teste da interface do Ampliar feito usando este método. Na seção seguinte descrevemos trechos dos testes e levantamos algumas das questões que surgiram durante este estudo em relação ao método utilizado. Finalmente, nas conclusões discutimos a relevância dos resultados para a pesquisa de IHC voltada ao domínio educacional.

SISTEMA AMPLIAR
O sistema Ampliar [6] é um sistema de apoio ao aprendizado desenvolvido no Instituto de Matemática e Estatística da Universidade do Estado do Rio de Janeiro (UERJ) para apoiar o ensino do método simplex no curso de Programação Linear. Na UERJ, o método simplex normalmente é ensinado aos alunos em 2 ou 3 aulas, nas quais o professor explica o método, exemplifica o seu uso e passa exercícios para os alunos. No entanto, a resolução de problemas pelo simplex implica a resolução de um grande número de operações algébricas. Estas operações consomem tempo, e requerem uma grande atenção, pois erros podem ser propagados e ter um alto custo para serem corrigidos. Assim, o aluno muitas vezes se sente desmotivado a fazer exercícios para fixar o conteúdo aprendido ou pode priorizar as operações em detrimento do entendimento do método. Além disso, os exercícios usualmente são feitos em horário extra-classe pelos alunos, o que dificulta o esclarecimento de dúvidas com o professor.

Com o objetivo de auxiliar os alunos na resolução de problemas pelo simplex e possibilitar um melhor entendimento do método é que o sistema Ampliar foi proposto e desenvolvido. Para que se conseguisse atingir no Ampliar os objetivos educacionais e de design desejados, ele foi desenvolvido dentro da abordagem da teoria da Engenharia Semiótica [1].

No design do Ampliar buscou-se dar espaço para os alunos tomarem as decisões sobre o passo a ser executado a cada momento. A cada passo, o aluno é quem deve solicitar ao sistema os cálculos das operações necessárias, fazendo assim com que ele tenha que saber os cálculos a serem executados, mas não cometa erros ao executá-los. O sistema apresenta sempre todas as opções ao aluno, mesmo aquelas que não estão corretas, para que o aluno decida e pense sobre o método. Quando o aluno comete um erro o sistema não lhe oferece de imediato uma

VI Simpósio sobre Fatores Humanos em Sistemas Computacionais — Mediando e Transformando o Cotidiano. Hotel Bourbon, Curitiba, 17 a 20 de outubro de 2004. UFPR, CEIHC—SBC.

explicação completa sobre o que ele deveria fazer, mas vai fornecendo gradativamente mais informações sobre o erro que está sendo cometido.

O sistema pretende não apenas permitir ao aluno automatizar a execução do método, mas também possibilitar seu entendimento. Assim, o sistema permite ao aluno tirar suas dúvidas sobre o método durante a execução do problema. Para isso, a cada passo são oferecidas ao aluno diversas perguntas e respostas organizadas nas categorias: "O que é?", "Como?", "E agora?", "Quando?" e "Por que?". Para cada categoria um conjunto de perguntas relevantes à etapa do método que ele se encontra são apresentadas, permitindo assim ao aluno um melhor entendimento do que significam os elementos envolvidos no método, como executar cada etapa ou sub-etapa, o que fazer a cada momento, quando executar cada cálculo e por que uma etapa ou cálculo é necessário ou relevante.

DESCRIÇÃO DO TESTE COM USUÁRIOS

Os testes de avaliação foram feitos no laboratório do SERG [9], utilizando o método de avaliação de comunicabilidade e contou com a presença de 7 alunos do curso de Programação Linear da UERJ.

Método de Avaliação de Comunicabilidade

O método de avaliação de comunicabilidade [4, 5] tem como objetivo avaliar uma interface com relação à qualidade da comunicação do designer para os usuários. Para isto, este método simula a comunicação do usuário para o designer sobre os problemas de interação que o usuário teve ao utilizar a interface. Isto é feito através de um pequeno conjunto de expressões que o usuário potencialmente pode usar para se exprimir em uma situação onde acontece uma ruptura na sua comunicação com o sistema. A aplicação do método pode ser dividida em duas etapas: a coleta de dados e a sua análise.

A coleta de dados é feita através de um teste em laboratório, durante o qual se grava um filme da interação, usando para isto um software de captura das ações do usuário. No caso do Ampliar, o filme da interação foi complementado com anotações dos observadores durante o teste, e uma entrevista pós-teste com cada aluno.

Analisa-se então o filme da interação gerado, utilizando as anotações para facilitar o entendimento do contexto e possíveis linhas de raciocínio dos participantes do teste. A análise é feita em 3 passos: (1) etiquetagem na qual se atribui a cada momento de ruptura da interação observado no filme a expressão apropriada (a partir do conjunto proposto). Por exemplo, quando o usuário percorre opções do menu à procura de uma determinada função, então se associaria a esta porção da interação a expressão "Cadê?"; (2) a interpretação das expressões associadas através da sua associação a ontologias de problemas de interação ou diretrizes de design; e (3) um perfil semiótico, na qual, a partir dos passos anteriores e com base no quadro teórico da Engenharia Semiótica, tenta-se reconstruir a meta-mensagem do designer para o usuário através da interface.

Perfil dos Alunos

Participaram da avaliação do Sistema Ampliar 5 homens e 2 mulheres, alunos do 5º ao 8º período dos Cursos de Matemática, de Informática ou de Engenharia Mecânica da UERJ. Em sua maioria, os alunos participantes utilizam o computador diariamente, em média por 30 a 40 horas semanais, acessando-o tanto do trabalho como de casa.

Todos os participantes pertenciam a uma mesma turma da disciplina Programação Linear. Por ocasião da avaliação, os alunos haviam assistido às aulas sobre o método simplex e – idealmente – deveriam estar executando uma lista de exercícios para consolidar seu aprendizado. Dado que ainda estariam na etapa de consolidação de conhecimentos sobre o simplex, os alunos desta turma mostraram-se especialmente representativos do perfil de usuários-alvo para o qual o sistema Ampliar foi desenvolvido. Os alunos recrutados voluntariamente se interessaram em participar da avaliação de comunicabilidade do Ampliar.

Teste Aplicado

Durante a aplicação do teste foi apresentado a cada aluno uma breve descrição do sistema Ampliar. Nesta descrição foi explicado o objetivo do Ampliar, ressaltando a possibilidade de se fazer perguntas sobre o conteúdo ao próprio sistema e a apresentação gradativa da explicação completa no caso de erro. Além disso, foram apresentadas algumas particularidades da interface, como a necessidade de se dar *scroll* para se visualizar os cálculos efetuados e a representação da notação de alguns elementos na interface.

Após o aluno ter confirmado o entendimento do texto explicativo, ele recebeu então um exercício para resolver utilizando o Ampliar. O exercício passado era pequeno, de solução simples e semelhante a um exercício apresentado em sala de aula.

Os testes foram aplicados por uma especialista em IHC, e observados por uma professora de Programação Linear.

DESAFIOS NA AVALIAÇÃO

A aplicação do teste com os alunos e a coleta dos dados pôde ser considerada habitual, no sentido de que não exigiu cuidados diferentes daqueles que se têm na avaliação da comunicabilidade da interface de qualquer outro software. No entanto, ao iniciarmos a análise, nos deparamos com algumas situações de ruptura de interação que deveriam ter um tratamento distinto do proposto no método original. Vale ressaltar que a análise foi feita por 2 especialistas no método de avaliação de comunicabilidade (EC) e 1 especialista no domínio (ED).

A partir da análise, classificamos estas situações em 3 categorias de acordo com o tratamento diferenciado no método que consideramos necessário: (1) ambigüidade de sintoma – a expressão associada aos sintomas observados não descreve a ruptura ocorrida; (2) novos sintomas – ações do usuário que normalmente não eram um sintoma de ruptura, deveriam neste contexto ser encaradas como tal; (3) alterações na interpretação da ruptura observada. Nesta seção apresentamos para cada uma das categorias identificadas as descrições do que foi observado no teste, as questões envolvidas na análise e discutimos os pontos sobre a avaliação da interfaces educacionais que elas levantam.

Ambigüidade de Sintoma

Observação no Teste
Durante o teste do Participante 7 (P7) verificamos que em um determinado momento, após executar alguns cálculos, seu cursor é colocado sobre a área em que estes cálculos são mostrados e fica por um tempo parado ou se alternando entre dois trechos dos cálculos apresentados.

Questão
Normalmente no método de avaliação de comunicabilidade este comportamento durante a interação seria etiquetado com um "E agora?" indicando que o usuário não sabia como prosseguir. No entanto, neste ambiente como o cursor estava parado sobre uma a área de cálculos as avaliadoras EC não souberam identificar se o aluno não sabia como prosseguir, o que seria uma ruptura de interação, ou simplesmente se estava raciocinando sobre o que o sistema acabara de lhe apresentar. Na avaliação da ED fica claro que era o 2o. caso, uma vez que ela identificou que os trechos de cálculo sobre os quais o cursor se alternava eram justamente aqueles necessários naquele momento para se decidir o próximo passo na resolução do problema.

Discussão
Na avaliação do Ampliar ficou claro que eram necessárias a presença de um especialista do método de avaliação de comunicabilidade e outro do domínio para fazer uma análise precisa da interação. Assim, o ponto levantado é qual o conhecimento necessário para se fazer a avaliação de um teste com usuários de um ambiente educacional. São necessários especialistas distintos ou poderia ser dado a um deles subsídios para tirar este tipo de ambigüidades (ou algumas delas)? Que subsídios seriam estes? Qual seria o custo x benefício de se ter todos os especialistas necessários ou de se abrir mão de algum deles?

Novos Sintomas

Observação 1 no Teste
O Participante 3 (P3) em um determinado ponto da resolução do problema não sabe como continuar. Apesar de fazer algumas perguntas ao sistema, ele continua sem entender o que deve fazer para tomar as decisões necessárias naquele momento. Após um tempo, o usuário simplesmente desiste de entender e continua a resolução do problema na base da tentativa e erro.

Questão 1
Neste caso, o que foi observado não caracteriza uma ruptura da comunicação sobre a interface, mas sim uma ruptura sobre a comunicação sobre o método simplex. No entanto, o método de avaliação de comunicabilidade não permite aos avaliadores o registro deste tipo de ruptura. Vale ressaltar que a indicação desta ruptura não é baseada em uma ação do usuário, mas em um conjunto de ações que incluem muitas vezes outras rupturas.

Observação 2 no Teste
Nos testes dos participantes P3 e P4, em alguns momentos em que tinham que tomar alguma decisão, sua primeira iniciativa era sempre fazer todos os cálculos disponíveis, mesmo aqueles que não eram necessários para sua decisão.

Questão 2
Neste caso as ações não representam uma ruptura na comunicação sobre a interface. O aluno pode sempre que desejar (e achar necessário) fazer os cálculos. Até porque um dos cálculos disponibilizados não é necessário nunca para a resolução do problema, ele só é útil na verificação da melhoria na solução a cada iteração do método. No entanto, aqui existe uma falha da interface à medida que ela perde a oportunidade de deixar claro para o aluno os cálculos necessários ou sua utilidade na solução do problema. Além disso, ela ainda pode estar encorajando um entendimento errôneo de que todos os cálculos são sempre necessários.

Discussão
Estes dois exemplos apontam para a necessidade de se identificar outros tipos de quebras da comunicação designer-usuário que são referentes não às opções de interação, mas ao aprendizado associado a elas. Uma investigação mais profunda seria necessária para se tentar identificar se existe um conjunto adicional de expressões que descrevam estas rupturas na comunicação referente ao processo de ensino e aprendizado, ou se novas expressões surgem no contexto específico do objetivo educacional de uma determinada aplicação.

O primeiro exemplo levanta ainda uma outra consideração a ser feita sobre a aplicação do teste. Sabe-se que, muitas vezes, participantes de teste, embora avisados de que o sistema é quem está sendo testado, se sentem também avaliados. Em um contexto educacional, principalmente o do Ampliar, no qual o aluno-usuário deve ter um conhecimento mínimo para conseguir usá-lo, o teste pode explicitar a falta do conhecimento esperado do aluno. Mesmo que esta exposição não tenha nenhuma conseqüência prática para o aluno em termos do seu curso, ela pode ser bastante desconfortável para ele, principalmente se o professor do tópico estiver presente durante o teste. Assim, formas de lidar com isso devem ser uma preocupação permanente na preparação e execução de um teste de um sistema educacional com alunos.

Novas Interpretações

Observação no Teste
O Participante 5 (P5) recorre às perguntas a todo momento, para entender o que são os elementos do problema e o que fazer. Fica claro que o aluno tem tantas dúvidas, que de fato sua tentativa era de aprender o método, e não consolidá-lo. Isto fica ainda mais patente no momento em que o aluno interrompe o teste e se diz incapaz de resolver o problema.

Questão
No método de avaliação de comunicabilidade a falha em executar a tarefa e desistência do uso de um sistema é marcado com a expressão "Não dá." Normalmente este é um problema bastante grave uma vez que indica que a solução do designer para o problema do usuário não o satisfez. No caso de ambientes educacionais, um "Não dá" pode não ser um ponto negativo na sua avaliação, e pode ser até positivo. No caso do Ampliar, ao desistir o aluno disse: "Descobri que estou totalmente perdido." Esta descoberta pode ter permitido ao aluno avaliar o seu nível de conhecimento sobre o Simplex contra o esperado, ou mesmo o de seus colegas e ter sido um estímulo para que ele o estudasse. Assim, pode-se dizer que o Ampliar atingiu seu objetivo ao comunicar

claramente ao aluno qual era o conhecimento esperado que ele tivesse.

Discussão

Este exemplo indica que pode ser necessário associar as expressões do método de avaliação de comunicabilidade a ontologias propostas ou adaptadas ao domínio educacional.

CONCLUSÃO

Neste trabalho apresentamos uma avaliação formativa de um sistema educacional utilizando-se o método de avaliação de comunicabilidade. A partir deste estudo de caso levantamos questões específicas ao uso deste método de avaliação para ambientes educacionais. Algumas destas questões parecem extrapolar o método de avaliação de comunicabilidade, e devem ser consideradas mesmo na utilização de outros métodos de avaliação.

Apesar dos desafios encontrados, a avaliação feita forneceu um feedback bastante produtivo para o design do Ampliar. Assim, a discussão apresentada tem por objetivo também fornecer a equipes de desenvolvimento de software educacional alguns parâmetros que lhes permita avaliar o custo x benefício de uma avaliação formativa utilizando o método de comunicabilidade, ou até mesmo tentar definir uma adaptação do método para seu contexto específico.

Este trabalho contribui também para a pesquisa do método de avaliação de comunicabilidade à medida que mostra que o método proposto pode demandar adaptações em domínios específicos, como no caso do educacional. Apenas com uma investigação mais detalhada poder-se-á discutir se uma proposta de adaptação para todo o domínio educacional seria possível e qual seria ela. Além disso, neste estudo de caso identificamos rupturas tanto na comunicação sobre as opções de interação, quanto sobre o conteúdo apresentado pela interface, e ainda sobre o conhecimento presumido do aluno. Para a avaliação formativa, em particular, seria interessante se poder identificar em qual destes aspectos desta comunicação ocorreu a quebra. Assim, seria interessante se refletir sobre se, e como, estes aspectos da comunicabilidade da mensagem do designer para o usuário neste domínio têm repercussões na prática (e.g. no método) e até mesmo na teoria.

AGRADECIMENTOS

Agradecemos aos alunos da UERJ que participaram do teste, à UERJ pelo seu apoio à pesquisa do Ampliar e ao SERG pelo seu apoio à avaliação com usuários.

REFERÊNCIAS

1. de Souza, C.S. *The Semiotic Engineering of Human-Computer Interaction*. The MIT Press, Cambridge, MA, in press, 2004.
2. Lewis, C., Brand, C., Cherry, G., Rader, C. Adapting User Interface Design Methods to the Design of Educational Activities. *Proc. CHI 98*, USA, pp 18-23.
3. Nielsen, J. Heuristic Evaluation,*Usability Inspection Methods*. NY: John Wiley & Sons, 1994, 25-62.
4. Prates, R.O.; de Souza, C.S.; Barbosa, S.D.J.; "A Method for Evaluating the Communicability of User Interfaces.". *Interactions* 7, 1. NY: ACM, 31-38, 2000.
5. Prates, R. O., Barbosa, S. D. J. Avaliação de Interfaces de Usuário - Conceitos e Métodos. JAI, SBC, 2003.
6. Prates, R. O.; Figueiredo, R. M. V. de; Bach, C. F. Um Modelo de Apoio ao Projeto de Design de Interfaces de Ambientes de Aprendizado. *Anais do WIE*, SBC, 2003. V.541- 552.
7. Preece, J.; Rogers, Y.; Sharp, E. *Interaction Design: Beyond HCI*, NY, John Wiley & Sons. 2002.
8. Shneiderman, B. Designing the User Interface, 3rd. Edition. Reading, MA: Addison Wesley, 1998.
9. SERG – Semiotic Engineering Research Group, http://www.serg.inf.puc-rio.br/
10. Squires, D.; Preece, J. "Predicting quality in educational software: Evaluating for learning, usability and the synergy between them". *Interacting with Computers*, 11, 1999, 467–483.