

Testes de Usabilidade: exigência supérflua ou necessidade?

Ana Amélia Amorim Carvalho
Departamento de Currículo e Tecnologia Educativa
Instituto de Educação e Psicologia
Universidade do Minho
aac@iep.uminho.pt

Resumo – Os testes de usabilidade têm vindo a ser realizados por todos os que desenvolvem software educativo. Contudo, erros de diversa ordem são frequentemente cometidos durante a sua aplicação.

Esta comunicação pretende alertar para alguns aspectos a ter em atenção quando se levam a cabo esses testes.

Começamos com uma definição do conceito de usabilidade, passando depois a referir os diferentes tipos de testes de usabilidade, quando devem ser realizados, a selecção dos diferentes tipos de avaliadores e as suas funções, os instrumentos para a recolha de dados. Por fim, mencionam-se ainda aspectos gerais e éticos da realização de testes com utilizadores.

A necessidade de realizar testes de usabilidade está razoavelmente interiorizada em todos aqueles que desenvolvem software, particularmente software educativo multimédia¹. Contudo, apercebemo-nos que muitos os encaram mais como uma exigência supérflua, não se dedicando a uma planificação meticulosa dos testes a realizar nem fazem uma selecção da amostra conveniente, utilizando não raras vezes a amostra que está mais "à mão". Por esse motivo, pareceu-nos ser imprescindível realizar uma pequena reflexão que nos ajude a compreender a verdadeira importância da realização destes testes. Nesse sentido, vamos mencionar os diferentes tipos de testes, os diferentes tipos de avaliadores e as suas funções, os instrumentos para a recolha de dados e os dados a recolher.

Feita esta breve introdução, vamo-nos debruçar sobre o conceito de usabilidade, que integra múltiplos componentes e aplica-se a todos os aspectos de um sistema com que o utilizador tem que interagir, incluindo a instalação e manutenção, como refere Nielsen (1993).

A usabilidade está directamente relacionada com a interface (Nielsen, 1990, 1993; Grudin, 1992; Shneiderman, 1992; Dix et al, 1993; Hix e Hartson, 1993; Preece et al, 1994; Baecker et al., 1995; Martins, 1995; Smith e Mayes, 1996), que juntamente com o utilizador e o sistema computacional interactivo constituem os três principais componentes da Interação Homem-Computador (Martins, 1995). Os primeiros estudos de usabilidade surgiram no

¹ O termo multimédia, neste texto, é entendido como um documento interactivo, que pode integrar texto, imagem, som e vídeo, em suporte digital, sendo, geralmente, a sua informação apresentada de forma não sequencial. Neste sentido, o termo multimédia é sinónimo do termo hipermédia (cf. Carvalho, 1999).

âmbito de investigação realizada na Interação Homem-Computador (IHC). A usabilidade tem vindo a impôr-se como área de importância crescente, sendo regulada por ISO /DIS 9241: 11. A usabilidade é uma qualidade que deve ser inerente ao documento e que possibilita que os utilizadores o usem com satisfação, eficácia e eficiência na realização de tarefas (Babo, 1996). Um software pode estar bem concebido em termos de funcionalidade, mas se a sua usabilidade não for boa, o utilizador rejeitá-lo-á.

Várias posições surgem em torno do conceito de usabilidade e sobre elas nos vamos debruçar. Por exemplo, Shackell (1986) considera quatro parâmetros para medir a usabilidade: eficiência, aprendizagem, flexibilidade e atitude do utilizador. Para Hix e Hartson (1993), a usabilidade está relacionada com a eficácia e a eficiência da interface e com a reacção do utilizador à interface.

Nielsen (1993; 1995) enumera cinco parâmetros, que considera como tradicionalmente aceites, para medir a usabilidade: *fácil de aprender*, o utilizador rapidamente consegue interagir com o sistema, aprendendo as opções de navegação e a funcionalidade dos botões; *eficiente para usar*, depois de ter aprendido como funciona, consegue localizar a informação que precisa; *fácil de lembrar*, mesmo para um utilizador que usa o sistema ocasionalmente, não tem necessidade de voltar a aprender como funciona, conseguindo lembrar-se; *pouco sujeita a erros*², os utilizadores não cometem muitos erros durante a utilização do sistema, ou se os cometem devem conseguir recuperar, não devendo ocorrer erros catastróficos; e *agradável de usar*, os utilizadores sentem-se satisfeitos com o sistema, gostam de interagir com ele. Para Smith e Mayes (1996), a usabilidade atenta basicamente em três aspectos, respectivamente, *facilidade de aprendizagem*, *facilidade de utilização* e *satisfação no uso do sistema pelo utilizador*.

Considerámos que a proposta de Smith e Mayes (1996), de certo modo, engloba as propostas dos outros autores mencionados e coincide com os parâmetros que identificámos como pertinentes para se avaliar a usabilidade de um documento interactivo. Assim, um documento multimédia para ser facilmente aceite pelo utilizador deve ser fácil de aprender a usar, deve ser fácil de utilizar e deve provocar satisfação no utilizador.

(i) **fácil de aprender**: é um dos atributos mais importantes da usabilidade e que pode levar a que os futuros utilizadores optem por usar determinado documento em detrimento de outro. Este atributo deve ser medido mesmo em relação a utilizadores com

² No texto original, a expressão usada é "some errors" (Nielsen, 1993). Posteriormente, em 1995, refere que como a maior parte dos hiperdocumentos, geralmente, não são usados para aplicações como medicina ou gestão financeira, essa cláusula sobre a prevenção de erros catastróficos deixa de ser tão importante.

pouca experiência, como refere Nielsen (1993). Usando utilizadores que nunca usaram o documento, mede-se o tempo que demora até o utilizador (com pouca ou nenhuma experiência e utilizadores experientes) navegar à vontade. O utilizador deve compreender com facilidade a interface, os diferentes percursos e o que pode fazer no documento. Para isso, as instruções ao utilizador devem ser concisas e claras e ir aparecendo à medida que o utilizador acede aos diferentes percursos. Apresentar uma longa descrição no início do documento sobre como interagir em todos os percursos, torna-se aborrecido e pouco eficiente, porque se expõe o utilizador a muita informação. Por outro lado, deve estar disponível o acesso à "ajuda" sempre que o utilizador a necessite. Tenha em atenção que "ajuda" em excesso irrita o utilizador experiente e a falta dela desorienta o utilizador inexperiente.

(ii) **Fácil de utilizar:** depois do utilizador ter aprendido a interagir com o documento, deve conseguir usá-lo com facilidade, mesmo quando usa o documento ocasionalmente. Um outro ponto a ter em atenção nesta dimensão prende-se com a orientação do utilizador no documento.

(iii) **Satisfação do utilizador:** o utilizador gosta de navegar no documento, devido à interface, ao conteúdo disponível, à estrutura do documento, ao processo de interacção e navegação, às ajudas disponíveis entre outros.

Vários testes têm sido desenvolvidos para medir a satisfação do utilizador, a título de exemplo, podemos mencionar o SUMI (Software Usability Measurement Inventory) e o QUIS (Questionnaire for User Interface Satisfaction). O QUIS, desenvolvido por Chin et al. (1988), apresenta quatro factores: aprendizagem, terminologia e fluxo da informação, output do sistema, características do sistema. É constituído por dezassete perguntas e usa uma escala de diferencial semântico. O SUMI foi desenvolvido no âmbito do projecto MUSIC (Metrics for Usability Standards in Computing) pelo Human Factors Research Group, em Cork, e é constituído por cinco factores: gosto, eficiência, aprendizagem, ajuda e controlo (KiraKowski, 1996; Macleod et al., 1997). Tem cinquenta itens e utiliza uma escala de likert de três pontos. Com base no SUMI foram criados também o MUMMS (Measuring the Usability of Multi-Media) para documentos multimédia e o WAMMI (Website Analysis and MeasureMent Inventory) para sites na web.

Nielsen (1995) alerta para a "usabilidade internacional", para a qual contribuem factores sociais³ e culturais. Este tipo de usabilidade é de particular relevância no caso de se adoptar o mesmo sistema informático ou software educativo em países culturalmente diferentes.

Quando se realizam os testes

Os testes de usabilidade devem ser feitos ao longo do processo de desenvolvimento, mesmo durante a fase de concepção da interface como referem, entre outros, Rubin (1994), Smith e Mayes (1996).

Rubin (1994) considera relevante fazer testes exploratórios e testes de avaliação numa fase inicial de desenvolvimento, enquanto que o teste de validação deve ser realizado numa fase adiantada do processo. O *teste exploratório* deve ser realizado numa fase inicial do processo de desenvolvimento, quando se está a definir e a conceber um serviço ou recurso. Há que inventar modos de descobrir como os utilizadores respondem a determinado serviço, podendo-se utilizar “workshops” de visualização. Para verificar e esclarecer o serviço que os utilizadores precisam ou querem, pode-se propor aos utilizadores determinado serviço e pedir-lhes para explorarem como o antevêm. Eles vão indicar o que faz falta e o observador pode ver como tal serviço pode afectar as suas actividades. Este teste também pode analisar se a interface representa bem classes de objectos, se as relações entre esses objectos são facilmente compreendidas, se as especificação de pré-requisitos para usar o documento ou o produto são necessárias e, ainda, se a tabela de conteúdos está organizada de tal forma que agrada a utilizadores iniciados e a utilizadores experientes.

Numa fase inicial de desenvolvimento do protótipo, os testes sobre os ícones podem ser feitos em papel, solicitando ao sujeito para mencionar o que representam (Nielsen, 1993; 1995), para se verificar se o ícone transmite a ideia pretendida. Numa fase mais avançada do documento, mas ainda em construção, pode-se apresentar a imagem de uma página do protótipo e apresentar uma breve descrição do que cada ícone pode fazer, solicitando ao sujeito para identificar o ícone com a descrição (cf. Carvalho, 1999). Os resultados destes testes medem-se através da proporção dos ícones correctamente descritos ou nomeados.

³ Nielsen (1995) considera que para a aceitação de um sistema contribuem a aceitação social e a aceitação prática que integra os custos, a compatibilidade com outros sistemas, a fiabilidade e a funcionalidade. Por sua vez, a funcionalidade integra a utilidade e a usabilidade.

Apresenta como exemplo de aceitação social o sistema LYRE que permite ensinar poesia, propondo diferentes interpretações. O aluno pode fazer as suas anotações, mas não pode adicionar novas leituras, ficando essa tarefa reservada ao professor. Em França foi bem aceite o LYRE, mas não o foi na Dinamarca porque consideraram que limitava as potencialidades do aluno para uma descoberta independente.

O *teste de avaliação* deve ser realizado numa fase inicial ou para expandir os resultados do teste exploratório. Pode-se constituir uma equipa multidisciplinar para executar determinado serviço e para discutir as questões de usabilidade. O utilizador realiza tarefas, em vez de percorrer o produto e comentar sobre os écrans, e podem ser recolhidos dados quantitativos.

O *teste de validação*, também referido como teste de verificação, é realizado numa fase adiantada do processo e pretende verificar a usabilidade do serviço e a eficácia dos recursos de aprendizagem. Ao contrário dos outros dois testes que podem ocorrer durante a fase de desenvolvimento, este só se realiza na fase final. Revê-se a consistência do sistema, a interface. Compara-se o sistema com os "standards" de usabilidade, com orientações gerais e com outros serviços relacionados.

Rubin (1994) sugere ainda o *teste de comparação* que pode ser realizado em qualquer fase do processo de desenvolvimento. Permite comparar dois ou mais "designs" alternativos. Por exemplo, analisar qual dos estilos de interface se aproxima mais do modelo conceptual da população alvo. Identificar em qual deles se sentirá o utilizador mais orientado e em que tarefas é necessário disponibilizar a "ajuda". Este teste pode ser usado em conjunção com qualquer um dos outros testes.

O que é avaliado

Segundo Rubin (1994) deve-se medir o desempenho e as preferências do utilizador. A usabilidade é medida em relação a determinado tipo de utilizadores e de tarefas (Nielsen, 1993).

A título de exemplo, podemos dizer que, para medir o desempenho dos sujeitos, se pode contabilizar o tempo para completar cada tarefa; o número e percentagem de tarefas completadas correctamente; o tempo gasto a ler determinada secção; os ícones seleccionados incorrectos, as visitas ao índice, à tabela de conteúdos; os "comentários negativos", entre outros (Rubin, 1994).

Para medir as preferências dos sujeitos, Rubin (1994) indica que se atente nas expectativas que os sujeitos têm sobre a utilidade do produto; a facilidade em o usar, em aprender e em o instalar.

Os avaliadores

Verificamos haver consenso sobre os dois tipos de avaliadores que devem proceder à avaliação da usabilidade do documento, por um lado, estes devem ser convidados *especialistas* em Interação Homem-Computador e em multimédia e, por outro lado, devem

ser convidados sujeitos provenientes da mesma população dos futuros utilizadores (Marchionini, 1990; Shneiderman, 1992; Nielsen, 1995; Smith e Mayes, 1996).

Por sua vez, Tessmer (1993) considera que um sistema deve ser avaliado por três tipos de avaliadores: especialista, utilizador e pequeno grupo de utilizadores. Deste modo, salienta o contributo que pode ser dado por um sujeito e por um grupo.

Um especialista revê o sistema - A avaliação é feita por especialistas em IHC e em multimédia que podem detectar facilmente problemas de inconsistência do sistema, tarefas pobres, interface confuso, entre outros. Por exemplo, verificar se a mesma informação surge sempre em determinada localização em todos os écrans de forma a facilitar o seu reconhecimento pelo utilizador, diminuindo a necessidade de memorização.

Teste um observador/um utilizador - O observador observa o utilizador e vai anotando as dificuldades daquele. Este tipo de recolha de dados pode ser afectado pelo efeito de Hawthorne, como apontam, entre outros, Marchionini (1990) e Preece et al. (1994), que consiste no facto de os utilizadores ao saberem que estão a ser observados alterarem, por influência da observação, o modo como executam as tarefas. Mayes et al. (1990b) sugerem a observação de pares de utilizadores, para se atentar nas suas interacções verbais, como eles fizeram para avaliar o 'Strath Tutor'. Outro método, por vezes utilizado, consiste em pedir ao utilizador para expressar verbalmente o que está a pensar ('think aloud').

Alguns inconvenientes da observação directa podem desaparecer caso se usem várias câmaras de vídeo para depois sincronizar dados relativos ao écran, ao movimento das mãos e às expressões faciais (Marchionini, 1990; Shneiderman, 1992; Nielsen, 1993; Preece et al., 1994; Rubin, 1994).

Pequeno grupo de utilizadores - O desempenho dos membros do grupo é comparado. Sugere-se que consulte o estudo realizado por Carvalho (1999).

Um dos aspectos que temos que ter em atenção consiste na *selecção da amostra* de utilizadores. Os utilizadores que vão constituir a amostra têm que ter as características dos futuros utilizadores. Nielsen (1990a) considera que há que analisar as diferenças individuais do utilizador, posteriormente, Nielsen (1993) apresenta o "cubo do utilizador" que caracteriza a experiência dos utilizadores relativamente a três dimensões: experiência com os computadores em geral, experiência com o sistema e conhecimento do domínio. A dimensão *experiência do sujeito com computadores* pode oscilar entre experiência mínima até intensa experiência. É importante verificar se o utilizador se sente à vontade, por exemplo, com o rato, com o "scrolling", com o Mac ou com o PC. A dimensão *experiência com o sistema*

pode oscilar entre um utilizador sem experiência até um utilizador experiente. O facto do utilizador se sentir à vontade para iniciar a navegação no documento vai depender, em parte, da sua experiência prévia com estes documentos. Por fim, a dimensão *conhecimento do domínio* pode oscilar entre ignorante e conhecedor do domínio. Aquele que tem alguma familiaridade com o domínio tem menos um obstáculo a combater na navegação num hiperdocumento.

Métodos de avaliação e técnicas de recolha de dados

Verifica-se uma certa uniformidade em aceitar quatro métodos: avaliação heurística, observação, sondagem e método experimental. A *avaliação heurística* é feita por peritos.

A *observação*, que é feita, geralmente, num laboratório, pode ser directa, mas aí pode ocorrer o efeito de Hawthorne já referido, ou indirecta, utilizando-se para o efeito câmaras de vídeo para gravarem a interacção do sujeito com o software educativo ou possibilitando ao observador uma situação em que observe sem ser visto. Durante essa interacção do sujeito com o documento multimédia pode-se proceder ao registo automático da mesma (“software logging”), através de um programa que é acoplado ao software multimédia. Pode-se, ainda, solicitar aos utilizadores para verbalizarem o que pensam (“think aloud”). Esta verbalização pode ser pedida a um utilizador, mas é pouco natural, ou a pares de utilizadores enquanto interagem. Ao utilizar o método de observação, é imprescindível que o observador use um guião, assim, sabe o que vai observar e o que se pretende que seja observado em particular. Dá mais trabalho antes da observação, mas depois facilita a recolha de dados e a análise dos mesmos.

Na *sondagem* (“survey”) podem ser utilizadas, basicamente, duas técnicas de recolha de dados: questionários ou entrevistas.

Por fim, o método experimental, que permite o controlo de variáveis.

Preece (1993) propõe ainda o método analítico ou formal que é aplicado numa fase inicial. Este método permite prever a usabilidade de uma interface antes de esta ser usada.

Planear o teste

Planear o teste é essencial. Comece por especificar a necessidade do teste, explicitando o que medir e o motivo pelo qual o vai fazer. Coloque as questões que devem ser respondidas. De seguida, defina o perfil do utilizador.

Selecione o método, indique as tarefas, defina os instrumentos a utilizar ou a desenvolver, de seguida, selecione a ou as técnicas de recolha de dados.

Especifique o equipamento necessário: tipo de computador e quantidade e software. Indique os critérios necessários para se considerar terminado o teste. Defina se é possível ou impossível a ajuda do investigador durante o teste, especifique os dados a serem coligidos e a forma como vão ser tratados (Nielsen, 1993; Rubin, 1994). Por último, resta a elaboração do relatório com as propostas de alteração a serem realizadas.

Aspectos gerais e éticos dos testes com utilizadores

Deve-se garantir o sigilo dos participantes, atribuindo-se-lhes um número de identificação. Tenha também presente que, por vezes, antes de proceder aos testes é necessário treinar os utilizadores relativamente a alguns aspectos da interface que não lhes são familiares como termos técnicos ou, mesmo, o uso do rato.

Antes de começar, tenha o cuidado de fazer com que o participante se sinta relaxado, para isso pode conversar sobre qualquer assunto nem que seja sobre o tempo, as férias ou o próximo feriado. Gaste o tempo necessário até que o utilizador se sinta calmo, zele pelas emoções e pelo bem-estar do utilizador. Informe-o que o que está a ser avaliado é o documento e não ele ou ela enquanto pessoa.

No final, não se esqueça de agradecer a colaboração. É também aconselhável mostrar disponibilidade para informar o participante sobre os resultados obtidos.

Conclusão

A usabilidade de um documento é uma condição essencial para a sua aceitação pelo utilizador. Por isso, todo o tempo gasto na observação e na conversa com os utilizadores é tempo bem gasto, que poderá contribuir para evitar surpresas desagradáveis e evitar a perda de muito tempo (e, às vezes, de dinheiro) na realização de futuras alterações. A utilização de testes de usabilidade numa investigação é não só útil como imprescindível, mas tenha bem presente a forma como realiza esses testes para não ocorrerem enviesamentos nos dados recolhidos, nem na selecção da amostra⁴.

Referências Bibliográficas

- Baecker, R., Grudin, J., Buxton, W. e Greenberg, S. (1995). *Human-Computer Interaction: Toward the Year 2000*. San Francisco, Ca: Morgan Kaufmann Publishers.
- Babo, R. M^a G. F. B. (1996). *A avaliação da Usabilidade de um Sistema*. Dissertação de Mestrado. Braga, Escola de Engenharia, Universidade do Minho.

⁴ Filhos, amigos ou vizinhos podem integrar a amostra, mas desde que tenham as características dos futuros utilizadores!

- Carvalho, Ana Amélia Amorim (1999). *Os Hipermedia em Contexto Educativo. Aplicação e validação da Teoria da Flexibilidade Cognitiva*. CEEP: Universidade do Minho.
- Grudin, J. (1992). Utility and usability: research issues and development contexts. *Interacting with Computers*, 4, 2, 209-217.
- Hix, D. e Hartson, H.R. (1993). *Developing User Interfaces: Ensuring Usability Through Product and Process*. New York: John Wiley & Sons.
- Kirakowski, J. (1996). The software usability measurement inventory: background and usage. P. Jordan, B. Thomas e B. Weedmeester (eds), *Usability Evaluation in Industry*. London: Taylor & Francis, 169-178.
- Macleod, M., Bowden, R., Bevan, N. e Curson, I (1997). The MUSIC performance measurement method. *Behaviour & Information Technology*. Vol.16, 4/5, 279-293.
- Marchionini, G. (1990). Evaluating Hypermedia-Based Learning. In D. H. Jonassen e H. Mandl (ed.), *Designing Hypermedia for Learning*. Berlin: Springer-Verlag, 355-373.
- Martins, F.M.J. (1995). *Métodos Formais na Concepção e Desenvolvimento de Sistemas Interactivos*. Tese de Doutoramento. Braga: Escola de Engenharia, Universidade do Minho.
- Mayes, T., Kibby, M. e Anderson, T. (1990). Learning About Learning from Hypertext. In D. H. Jonassen e H. Mandl (eds.), *Designing Hypermedia for Learning*. Berlin: Springer-Verlag, 227-250.
- Nielsen, J. (1990). Evaluating Hypertext Usability. In D. H. Jonassen e H. Mandl (eds), *Designing Hypermedia for Learning*. Berlin: Springer-Verlag, 147-168.
- Nielsen, J. (1993). *Usability Engineering*. New Jersey: Academic Press.
- Nielsen, J. (1995). *Multimedia and Hypertext: the Internet and beyond*. Boston: AP Professional.
- Nielsen, J. e Lyngbaek, U. (1990). Two field studies of hypermedia usability. In R. McAleese e C. Green (eds), *Hypertext: state of the Art*. Oxford: Intellect, 64-72.
- Preece, J. (1993). *A Guide to Usability: human factors in computing*. Addison Wesley, the Open University.
- Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. e Carey, T. (1994). *Human-Computer Interaction*. Workingham: Addison-Wesley Publishing.
- Rubin, J. (1994). *Handbook of Usability Testing*. New York: John Wiley and Sons.
- Shackell, B. (1986). Ergonomics in Design for Usability. E. Harrison e A. Monk (eds), *People and Computers: Designing for Usability. Proceedings of the Second Conference of the British Computer Society*. University of York, 44-64.
- Smith, C. e T. Mayes (1996). *Telematics Applications for Education and Training: Usability Guide*. Commission of the European Communities, DGXIII Project.
- Tessmer, M. (1993). *Planning and Conducting Formative Evaluations*. London: Kogan Page.