

**Cruz, Renato Rojas da ; Tiradentes, Virginia**

*Medidas de usabilidade : uma investigação da correlação entre instrumentos*

**Interaction South America (ISA 14). 6ta. Conferencia Latinoamericana de Diseño de Interacción, 2014**

Este documento está disponible en la Biblioteca Digital de la Universidad Católica Argentina, repositorio institucional desarrollado por la Biblioteca Central "San Benito Abad". Su objetivo es difundir y preservar la producción intelectual de la Institución.

La Biblioteca posee la autorización del autor para su divulgación en línea.

Cómo citar el documento:

Cruz, R.R. da, Tiradentes, V. Medidas de usabilidade : uma investigação da correlação entre instrumentos [en línea]. En: Interaction South America (ISA 14) : 6ta. Conferencia Lationamericana de Diseño de Interacción; 2014 nov 19-22; Buenos Aires : Interaction Design Association ; Asociación de Profesionales en Experiencia de Usuario ; Internet Society ; Universidad Católica Argentina. Disponible en:  
<http://bibliotecadigital.uca.edu.ar/ponencias/medidas-usabilidade-investigacao.pdf>

# Medidas de Usabilidade: Uma Investigação Da Correlação Entre Instrumentos

**Renato Rojas da Cruz**  
Universidade de Brasília  
design@renatorojas.com

**Virginia Tiradentes**  
Universidade de Brasília  
v.tiradentes@gmail.com

## RESUMO

O presente estudo propõe investigar medidas de elementos da usabilidade conforme definida na ISO 9241-11. Realizou-se um experimento que investiga a validade de três questionários – ASQ, PSSUQ e UMUX – além de confrontar tal resultado com dados de eficácia – medida por sucesso na tarefa; eficiência – medida por tempo na tarefa; e informações sociodemográficas. Realizou-se análise de correlação entre fatores a partir do escore que cada um dos 41 participantes atribuiu em cada uma das escalas. O resultado reforça a semelhança entre o conteúdo mensurado pelas três escalas. Pode-se observar também que o Escore do Sucesso na Tarefa possui uma correlação com o escore do ASQ, o escore geral do PSSUQ e com os fatores 1 e 2, que medem Utilidade do Sistema e Qualidade da Informação, respectivamente. A eficácia mostrou ter uma correlação mais forte com o fator de Utilidade do Sistema do PSSUQ porém a eficiência mostrou ser uma medida frágil nesse estudo, já que não obteve nenhuma forte correlação que pudesse ser generalizada.

## Palavras-chave do autor

HCI, Usabilidade, Medidas, Correlação.

## INTRODUÇÃO

Usabilidade é um conceito que deve ser analisado considerando diferentes contextos. Elementos como o ambiente, a tarefa a ser executada e as próprias características do usuário (suas habilidades, seus conhecimentos) são variáveis que tornam o estudo da usabilidade um tema complexo. O interesse de se mensurar a usabilidade em diferentes contextos vem da necessidade de se desenvolver produtos cada vez mais centrados o usuário. Nesse panorama, foram estruturadas disciplinas que estudam a relação do homem com os computadores.

A Interação Humano-Computador (ou Human-Computer Interaction – HCI) pode ser definida como uma disciplina preocupada com o design, avaliação e implementação de sistemas interativos computacionais para a utilização humana e com o estudo dos seus principais fenômenos

[12]. Destaca-se nessa definição a necessidade e a responsabilidade dos pesquisadores em avaliar sistemas interativos.

A preocupação em avaliar o design também está presente no conceito de Design Centrado no Usuário (ou *User-Centered Design* – UCD). A definição de UCD também cita a avaliação do design: “é a prática de princípios que envolvam ativamente o usuário para claro entendimento de requisitos dos usuários e das tarefas, para o design iterativo e sua avaliação e também para uma abordagem multidisciplinar” [11].

Outro conceito fundamental na área de HCI é a usabilidade. A definição mais frequentemente adotada vem da ISO 9241-11 de 1998. Ela diz que “usabilidade é a medida em que um produto pode ser usado por usuários específicos para alcançar objetivos específicos com eficácia, eficiência e satisfação em um contexto de uso específicos” [1].

Usabilidade também pode ser explicada a partir de outros fatores. Nielsen [5] coloca como sendo um atributo que explica quão fácil é o uso de interfaces a partir de cinco fatores: aprendibilidade, eficiência, memorabilidade, erros e satisfação. Já Tullis e Albert [13] consideram que usabilidade seja uma habilidade do usuário em utilizar um artefato para executar uma determinada tarefa com sucesso. Rubin [9] explica que quando um artefato é verdadeiramente usável o usuário é capaz de fazer o que ele quiser, da forma que ele espera ser possível fazer, sem resistência, hesitação ou dúvidas. Este conceito ajuda a explicar, entre outros fenômenos, a satisfação do usuário em utilizar um determinado sistema. Os resultados dos estudos em usabilidade tem guiado o desenvolvimento e a melhoria de inúmeros produtos pelo mundo. Nos estudos empíricos, o domínio de ferramentas que ajudem a apontar os caminhos de um produto com mais usabilidade torna-se cada dia mais relevante já que esse conceito tem sido utilizado para explicar o sucesso e o fracasso de vários produtos lançados no mercado.

A percepção de usabilidade é um traço latente e como tal não há forma direta de medida. Por isso torna-se fundamental a elaboração e uso de instrumentos que possam acessar e mensurar corretamente esse construto. O desenvolvimento e a validação de questionários para esse fim específico são uma forma eficiente, popular e de

baixo custo para se medir a percepção de usabilidade de um artefato digital.

Neste estudo uma análise comparativa de diferentes modos de medir a percepção de usabilidade atribuída a um website é feita com o objetivo de verificar a eficácia de tais métodos. Para tanto, foi realizado um experimento que visa investigar a validade de três questionários de usabilidade além de confrontar tal resultado com dados de eficácia – medida por sucesso na tarefa; eficiência – medida por tempo na tarefa; e informações sociodemográficas. O objetivo de utilizar procedimentos psicométricos para comparação dos dados é o de estabelecer um critério de comparação com outros instrumentos da área.

### QUESTIONÁRIOS PÓS-TAREFA

Desenvolver e validar questionários exige tempo e cuidado. Dentre os vários procedimentos utilizados para estudos de usabilidade, os questionários se destacam por possuírem um baixo custo de aplicação e fornecerem informação valiosa. Especificamente para o tipo de pesquisa em que o usuário é colocado em um cenário que reproduz condições reais de uso, existem os chamados Questionários Pós-Cenários (QPC). Os questionários escolhidos para esse estudo são o After-Scenario Questionnaire (ASQ), o Post Study System Usability Questionnaire (PSSUQ) e o Usability Metric for User Experience (UMUX). A decisão de utilizá-los levou em consideração, principalmente, a frequência com que esses instrumentos são citados pela literatura. O ASQ e o PSSUQ estão entre os instrumentos mais utilizados para medir usabilidade [13] enquanto o UMUX é uma síntese recente de outro instrumento amplamente utilizado, o System Usability Scale (SUS).

O ASQ nasceu de estudos realizados por James Lewis [6], pesquisador da IBM, para coletar dados sobre Satisfação do usuário. O questionário tem apenas três itens construídos em uma escala Likert de 7 pontos com âncoras no 1 como “Discordo fortemente”, no 7 como “Concordo fortemente” e “Não se aplica” como um ponto fora da escala. Cada item aborda um componente importante da Satisfação: Facilidade em Completar a Tarefa, Tempo para Completar a Tarefa e Adequação da Informação de Suporte. Concluiu-se a validade psicométrica do ASQ com um estudo conduzido com 48 participantes selecionados por conveniência de três cidades diferentes dos Estados Unidos.

O PSSUQ, assim como o ASQ, foi desenvolvido por Lewis. Elaborado para ser mais completo, o PSSUQ possui 19 itens construídos em uma escala Likert de 7 pontos com âncoras no 1 como “Discordo fortemente”, no 7 como “Concordo fortemente” e “Não se aplica” como um ponto fora da escala. A análise fatorial revelou a presença de três fatores: Utilidade do Sistema, Qualidade da Informação e Qualidade da Interface. A composição de cada fator está no Anexo A2.

O UMUX, desenvolvido por Finstad [10], pesquisador da Intel, surgiu de uma necessidade de redução do SUS, escala de 10 itens desenvolvida por Brooke [3]. O UMUX tem apenas quatro itens e é utilizado junto com outros instrumentos para medir a Experiência do Usuário (ou User Experience – UX), conceito mais amplo em relação a Usabilidade. A UX leva em consideração toda a interação com o artefato assim como pensamentos, sentimentos e percepções que resultam da interação [13]. O UMUX, que mostrou ter alta correlação com o SUS [10], se utiliza de uma escala Likert de 7 pontos ancorada no 1 com “Discordo fortemente” e no 7 com “Concordo fortemente”.

### MÉTODO

#### Design do experimento

Os participantes, estudantes universitários e colaboradores de uma agência de publicidade, foram convidados a navegar em um site governamental de notícias por cinco minutos onde teriam que realizar cinco tarefas que se julgou frequentes e de diferentes graus de dificuldade e complexidade.

Enquanto realizavam as tarefas, as imagens da tela foram gravadas por meio do software QuickTime Player versão 10.3. Após a realização das tarefas, foi solicitado aos participantes que respondessem um questionário único contendo três instrumentos – o ASQ, o PSSUQ e o UMUX. Esse procedimento torna possível comparar os resultados entre os instrumentos pois minimiza o efeito de qualquer outra variável.

Foi realizado um teste piloto com três participantes para verificar a dificuldade das tarefas. O resultado levou a pequenos ajustes na redação. Na primeira tarefa proposta ao participante, lia-se “Qual é a hora de publicação da principal notícia?”. Foi percebido que o termo “principal notícia” deixava dúvidas nos participantes e por isso este termo foi alterado para “notícia em destaque”. Na terceira tarefa proposta, lia-se “Qual é a hora de publicação da última notícia do Distrito Federal?”. Foi percebido que o termo “última” criava a dúvida sobre que notícia estava-se se referindo: se a mais recente ou se a mais antiga, dependendo de como se ordena a cronologia das notícias.

1	Qual é hora de publicação da notícia em destaque?
2	Quantos itens estão agrupados em Centrais de Conteúdo?
3	Qual é a hora de publicação da notícia do Distrito Federal mais recente?
4	Em que cor fica o fundo do site quando em alto contraste?
5	Em quantas redes sociais o Ministério do Turismo está presente?

**Tabela 1. Lista revisada de tarefas aos participantes.**

O termo “última notícia” foi assim alterado para “notícia mais recente”. A Tabela 1 mostra a lista revisada de tarefas aos participantes.

### Participantes

A pesquisa contou com um total de 41 participantes, com idades entre 17 e 62 anos. A média de idade é de 27 anos e desvio-padrão de 8,6 anos. A maioria é do sexo feminino (68,3%), com nível superior de escolaridade (63,4%) e afirma não ter navegado no site anteriormente ou não se lembrar se navegou anteriormente (75,6%). Gráficos 1 e 2 mostram o número de participantes por idade e por escolaridade respectivamente.

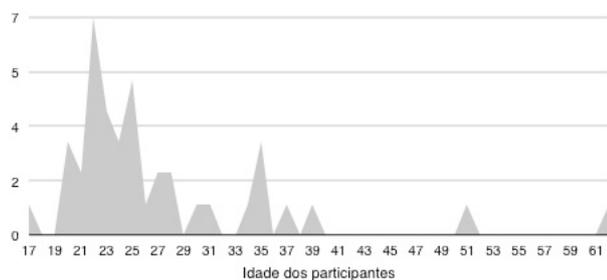


Gráfico 1. Número de participantes por idade.

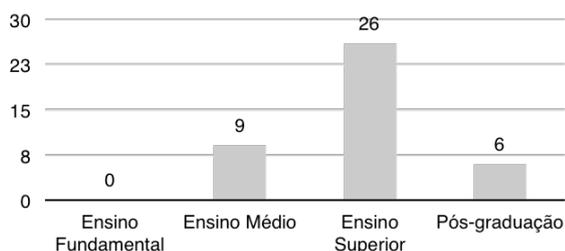


Gráfico 2. Número de participantes por escolaridade.

### Material

Para a navegação no site, os participantes utilizaram computador com conexão de banda larga e tiveram a opção de utilizar o trackpad de um laptop ou um mouse. O mesmo equipamento foi disponibilizado para que os participantes respondessem o formulário digital hospedado na plataforma GoogleDocs.

O questionário respondido pelos participantes continha 26 afirmações, composto de 3 escalas (o ASQ, o PSSUQ e o UMUX) do tipo Likert de 7 pontos. O ponto 1 estava ancorado com “Discordo fortemente” e o 7 com “Concordo fortemente”. Os questionários foram aplicados em ordem balanceada de forma que os três instrumentos fossem apresentados o mesmo número de vezes em cada uma das posições possíveis. Dessa forma, a versão A

apresentava o ASQ primeiro, o PSSUQ em seguida e finalmente o UMUX.

Para todas as perguntas o participante deveria utilizar uma escala de concordância de 7 pontos em que 1 estava ancorado com “Discordo totalmente” e 7 com “Concordo totalmente”. Caso o participante julgasse que a sentença não tinha conteúdo pertinente para avaliar a sua experiência em navegar no site, foi dada orientação para que ele deixasse-a em branco.

Ao final do questionário, o participante foi indicado a registrar seus dados pessoais como, sexo, idade, escolaridade, profissão e se havia alguma experiência prévia em utilizar o portal em questão. Toda a aplicação envolveu exclusivamente material em língua portuguesa.

### Procedimentos

A coleta de dados ocorreu no ambiente acadêmico e profissional. Nas duas situações foram tomadas medidas para que não houvesse interrupções durante a execução da tarefa. Em todos os casos, a conexão de internet permaneceu estável, sem que fosse um elemento impeditivo para o bom andamento da atividade.

Inicialmente o participante foi convidado e esclarecido sobre os objetivos da pesquisa por meio da apresentação do Termo de Consentimento Livre e Esclarecido. Em seguida, foi orientado sobre as tarefas que deveria realizar durante o tempo de cinco minutos de navegação no site. Foram registrados o Tempo de Tarefa e o Sucesso de Tarefa. Ao final, o participante foi conduzido para o formulário eletrônico contendo os instrumentos de usabilidade e o questionário sociodemográfico.

### RESULTADOS E DISCUSSÃO

No tratamento dos dados, foi calculado o escore que cada participante atribuiu em cada uma das escalas. Para tanto, partiu-se das estruturas fatoriais descritas pelos autores das escalas pois a amostra não foi robusta o suficiente para gerar uma matriz fatorável [2]. Sendo assim, o ASQ e o UMUX não foram considerados instrumentos fatoráveis por serem constituídos de um número reduzido de itens. Já o PSSUQ foi dividido em 3 fatores, como estudos anteriores já identificaram: Utilidade do Sistema, Qualidade da Informação e Qualidade da Interface.

Chamou a atenção a porcentagem de determinados itens deixados em branco. Observa-se um alto grau de abstenção nos itens que remetem à satisfação do usuário com suporte, como por exemplo o item “eu fiquei satisfeito com a informação de suporte (ajuda online, mensagens, documentação) ao completar tarefas”, “sempre que eu cometi um erro usando o site, pude recuperar-me fácil e rapidamente” e “as informações (tais como ajuda online, mensagens na tela e documentação) fornecidas pelo site são claras”. Este resultado pode indicar que o usuário não percebeu a presença de elementos específicos de suporte.

Na análise relativa às tarefas, foi observado o sucesso na resposta e o tempo levado por cada usuário para respondê-la. Pode-se considerar que a porcentagem de acerto de uma tarefa corresponde a seu indicador de dificuldade. Sendo assim, as Tarefas 1 (92% de acerto) e 2 (80% de acerto) são consideradas de baixa dificuldade enquanto a Tarefa 3 (17% de acerto) e 5 (0% de acerto) são consideradas de alta dificuldade. A tarefa 5 não teve nenhuma resposta correta observada. A tarefa 4 (45% de acerto) foi considerada de média dificuldade.

Foi calculado um escore geral para os questionários, para os três fatores do PSSUQ e para o Sucesso na Tarefa que variam de zero a um. Não foi calculado escore geral de tempo na tarefa pois quase que a totalidade dos participantes utilizaram todo o tempo disponível para realizar as tarefas. No caso dos itens 2 e 4 do UMUX, em que sua redação de é de caráter negativo, a pontuação na escala foi invertida. As médias estão apresentadas no gráfico 3. Pode-se afirmar que o UMUX tem a tendência de fornecer um escore levemente mais positivo que o ASQ e o PSSUQ. O escore do sucesso na tarefa não demonstrou ser uma medida relevante, já que varia muito de acordo com o tipo de tarefa apresentada ao participante. Isso não desclassifica o procedimento de em si pois foi essencial para guiar a navegação do participante no teste.

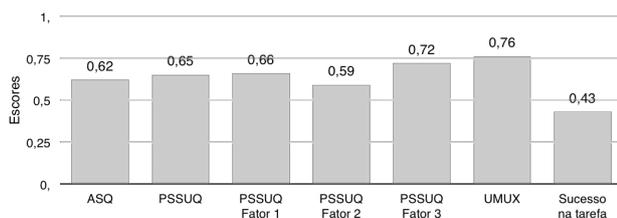


Gráfico 3. Escores das medidas e seus fatores.

Foi realizada uma análise de correlação entre fatores, que visa esclarecer o grau de associação entre duas variáveis [4]. Percebe-se que existe uma correlação de alta magnitude e significância entre o escore do ASQ, o escore do PSSUQ e seus três fatores e o escore do UMUX. Esse resultado reforça a semelhança entre o

conteúdo mensurado pelas três escalas (Tabela 2). Pode-se observar ainda que o Escore do Sucesso na Tarefa possui uma correlação moderada com o escore do ASQ, o escore geral do PSSUQ e com os fatores 1 e 2, que medem Utilidade do Sistema e Qualidade da Informação, respectivamente. Não foi encontrada correlação com o fator 3 do PSSUQ que mede Qualidade da Interface e com o escore do UMUX. Acredita-se que esse resultado é coerente já que o fator Qualidade da Interface aborda maioritariamente a satisfação do participante com a interface. A falta de correlação entre o Sucesso na Tarefa e o escore do UMUX aponta para a possibilidade dessas duas medidas serem complementares.

Foi encontrada correlação moderada entre alguns escores e os tempos na tarefa 1 e 4, conforme detalhado na Tabela 3. Os valores negativos de correlação indicam que quanto menor o tempo em que o participante executou a tarefa, maior a sua percepção de usabilidade. As tarefas 1 e 4 são as que, em média, foram respondidas mais rapidamente. A fraca correlação com o fator 3 do PSSUQ indica que a percepção da qualidade da interface medida pelo instrumento não está relacionada ao tempo em que o participante executou as tarefas.

Como esperado, não há correlação entre os escores e idade, tampouco entre os escore e escolaridade.

## CONCLUSÃO

O estudo aponta para a forte correlação entre os três instrumentos e a semelhança entre o conteúdo mensurado pelas três escalas. O PSSUQ leva desvantagem em relação ao número de itens, já que é muito mais extenso que os outros. Cabe ressaltar, entretanto, que o PSSUQ possui maior correlação com o indicador de eficácia (escore de sucesso na tarefa) e que o UMUX não apresentou correlação significativa com essa medida. Outra ressalva que deve ser feita diz respeito ao índice de abstenção observado nos instrumentos. Embora seja um instrumento sedutor por ser sucinto, um dos itens do ASQ chegou a apresentar 24,4% de respostas em branco. Isso torna o instrumento menos atraente já que aponta para o fato de que seu conteúdo pode não ser aplicável a qualquer website.

	Escore ASQ	Escore PSSUQ	Escore PSSUQ - Fator 1	Escore PSSUQ - Fator 2	Escore PSSUQ - Fator 3	Escore UMUX	Escore de Sucesso na Tarefa
Escore ASQ	1						
Escore PSSUQ	0,892**	1					
Escore PSSUQ - Fator 1	0,889**	0,979**	1				
Escore PSSUQ - Fator 2	0,884**	0,951**	0,901**	1			
Escore PSSUQ - Fator 3	0,719**	0,884**	0,818**	0,794**	1		
Escore UMUX	0,853**	0,899**	0,884**	0,855**	0,788**	1	
Escore de Sucesso na Tarefa	0,354*	0,400*	0,426**	0,383*	0,220	0,278	1

Tabela 2. Correlação entre os escores.

\*  $p < 0,05$  \*\*  $p < 0,01$

	Tempo na tarefa 1	Tempo na tarefa 2	Tempo na tarefa 3	Tempo na tarefa 4	Tempo na tarefa 5
Escore ASQ	-0,323*	-0,126	-0,163	-0,346*	-0,159
Escore PSSUQ	-0,325*	-0,073	-0,078	-0,333*	-0,099
Escore PSSUQ - Fator 1	-0,315*	-0,057	-0,109	-0,385*	-0,148
Escore PSSUQ - Fator 2	-0,349*	-0,145	-0,066	-0,312*	-0,100
Escore PSSUQ - Fator 3	-0,242	-0,010	-0,030	-0,163	0,012
Escore UMUX	-0,320*	-0,085	-0,123	-0,388*	-0,213
Escore de Sucesso na Tarefa	-0,286	-0,138	-0,043	-0,269	-0,048

Tabela 3. Correlação entre os escores e o tempo em cada tarefa.

\*  $p < 0,05$  \*\*  $p < 0,01$

A eficácia mostrou ter uma correlação mais forte com o fator de Utilidade do Sistema do PSSUQ porém a eficiência (medida pelo tempo da tarefa) mostrou ser uma medida frágil nesse estudo, já que não obteve nenhuma forte correlação que pudesse ser generalizada.

Sendo assim, foi possível apontar diferenças entre os instrumentos mas não o suficiente para apontar um instrumento como sendo melhor que o outro. Cada instrumento tem vantagens e desvantagens e isso deve ser levado em consideração de acordo com cada caso de pesquisa.

O volume da amostra utilizada forneceu um resultado robusto para o experimento. Ainda assim, sugere-se a replicação desse estudo com amostras maiores para permitir processos metodológicos para inspeção mais profunda da qualidade das medidas.

## REFERÊNCIAS

1. A. Abran, A. Khelifi, and W. Suryn, "Usability Meanings and Interpretations in ISO Standards," *Softw. Qual. J.*, vol. 11, pp. 325–338, 2003.
2. B.G. Tabachnik e L.S. Fidell, "Using Multivariate Statistics", 4ª ed. Harper Collins, New York.
3. J. Brooke, "SUS: A 'quick and dirty' usability scale", *Usability Evaluation in Industry*, Taylor and Francis, London, 1996.
4. J. Miles e M. Shevling, "Applying Regression and Correlation: A Guide for Students and Researchers", 2001.
5. J. Nielsen, "Usability 101: Introduction to Usability," 2012. [Online]. Disponível em: <http://www.nngroup.com/articles/usability-101-introduction-to-usability/>.
6. J. R. Lewis, "Psychometric Evaluation of an After-Scenario Questionnaire for Computer Usability Studies: the ASQ," *SIGCHI Bull.*, vol. 23, no. 1, pp. 78–81, 1991.
7. J. R. Lewis, "IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use," Janeiro, 1995.
8. J. R. Lewis, "Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies," *Int. J. Hum. Comput. Interact.*, vol. 14, no. 3, pp. 463–488, Sep. 2002.
9. J. Rubin, "Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests", segunda edição, 2008.
10. K. Finstad, "The Usability Metric for User Experience", *Interacting with Computers* 22 (5), Elsevier B.V., 2010.
11. K. Vredenburg, J.-Y. Mao, P. W. Smith, and T. Carey, "A Survey of User-Centered Design Practice," in *CHI*, 2002, no. 1, pp. 471–478.
12. R. Harper, T. Rodden, Y. Rogers, and A. Sellen, Eds., *Being Human – Human-Computer Interaction in the Year 2020*. 2008, p. 100.
13. T. Tullis e W. Albert, "Measuring the user experience: collecting, analyzing, and presenting usability metrics", segunda edição, 2013.