

Gresse von Wangenheim, Christiane ; Ferreti Borgatto, Adriano ; Vargas Nunes, Juliane ; Cardoso Lacerda, Thaisa ; Oliveira, Rafael José de ; Krone, Caroline ; Reolon, Mathias ; Xafranski, Jessica

Sure : uma proposta de questionário e escala para avaliar a usabilidade de aplicações para smartphones pós-teste de usabilidade

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:

Gresse von Wangenheim, Ch., Ferreti Borgatto, A., Vargas Nunes, J. et al. Sure : uma proposta de questionário e escala para avaliar a usabilidade de aplicações para smartphones pós-teste de usabilidade [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/sure-proposta-questionario-escala.pdf>

SURE: uma proposta de questionário e escala para avaliar a usabilidade de aplicações para smartphones pós-teste de usabilidade

Christiane Gresse Von Wangenheim¹, Adriano Ferreti Borgatto¹, Juliane Vargas Nunes², Thaisa Cardoso Lacerda¹, Rafael José de Oliveira, Caroline Krone¹, Mathias Reolon¹, Jessica Xafranski¹

¹Departamento de Informática e Estatística – Universidade Federal de Santa Catarina

²Departamento de Design e Expressão Gráfica – Universidade Federal de Santa Catarina

{adriano.borgatto, c.wangenheim} @ufsc.br, {julivn, tha.c.lacerda, faeljoliveira, kronecaroline, mathiasreolon, jessicaxafranski} @gmail.com

RESUMO

Os smartphones estão se tornando o dispositivo de uso pessoal mais popular e difundido. Com um modelo de interação diferente do tradicional para desktop, eles propõem novos desafios ao design de interfaces. Diante disso, a usabilidade tem sido considerada um dos mais importantes fatores de qualidade de aplicações para smartphone. Atualmente, existem vários questionários padronizados pós-teste de usabilidade. No entanto, eles variam quanto aos fatores de qualidade medidos, as escalas de resposta e são essencialmente voltados para computadores. Neste contexto, este artigo apresenta a concepção e validação de um questionário pós-teste, bem como uma escala para medir a usabilidade de aplicações smartphone. A partir de uma revisão sistemática da literatura e considerando características específicas de aplicações smartphone, foi desenvolvido um questionário inicial com 37 itens. Em seguida, foi realizado um estudo empírico com a aplicação desse questionário em 305 testes com usuários. Os resultados foram analisados utilizando a Teoria de Resposta do Item (TRI). Com base nessa análise foi concebida a versão atual do questionário, chamada de SURE (Smartphone Usability questionnaiRE), além de serem criadas uma versão resumida deste e uma escala de usabilidade.

Palavras-chave

Usabilidade, avaliação, questionário pós-teste de usabilidade, aplicações smartphone, teoria da resposta ao item.

INTRODUÇÃO

Os smartphones estão se tornando o dispositivo de uso pessoal mais popular e difundido (Accenture, 2012). Seu conceito de mobilidade gerou o surgimento de diversas

aplicações móveis (GASIMOV, 2010). As especificidades deste dispositivo têm um significativo impacto na natureza da interação (que pode utilizar gestos, sensores, câmera, voz, etc) o que propõe novos desafios para o design de interfaces (WASSERMAN, 2010).

A usabilidade é considerada um dos mais importantes atributos de qualidade das aplicações para smartphone, afetando a satisfação dos usuários (KETOLA e RÖYKKEE, 2001; TREERATANAPON, 2012; SEFFAH et al., 2006). Um dos principais métodos de avaliação de usabilidade sob a perspectiva do usuário é o teste de usabilidade, que envolve representantes do público-alvo realizando tarefas típicas no sistema. Contudo, a usabilidade percebida pelo usuário envolve uma série de aspectos subjetivos que não são diretamente observáveis durante o teste.

Uma forma de medir esses aspectos subjetivos é através da aplicação de um questionário no final do teste. Existem vários questionários pós-teste padronizados como SUS - System Usability Scale (BROOKE, 1996), QUIS - Questionnaire for User Interaction Satisfaction (CHIN, 1988), SUMI - Software Usability Measurement Inventory (KIRAKOWSKI e CORBETT, 1988) ou UMUX-LITE - Usability Metric for User Experience (LEWIS, 2013). Todos são ferramentas reconhecidas que foram submetidas à validação psicométrica, incluindo confiabilidade, validade e sensibilidade.

No entanto esses questionários variam muito quanto aos fatores medidos (apenas satisfação ou também a percepção subjetiva de eficácia e eficiência), número de itens, escalas de resposta, etc (HORNBEK, 2006). Além disso, tais questionários são voltados, principalmente, para medir a usabilidade de sistemas para computadores desktop. E, embora os atributos de usabilidade de aplicações desktop sejam considerados válidos para aplicações móveis (NIELSEN, 2014) existem outros aspectos que precisam ser considerados.

Assim, este artigo apresenta a concepção sistemática de um questionário pós-teste de usabilidade, bem como sua

validação a partir dos resultados de 305 testes com usuários, analisados estatisticamente com a aplicação da Teoria de Resposta ao Item – TRI e a criação de uma escala de medição de usabilidade.

PROCEDIMENTOS METODOLÓGICOS

Nessa pesquisa foi adotada uma metodologia baseada em Schwab (1980), dividida em três etapas, conforme ilustrado na Figura 1.

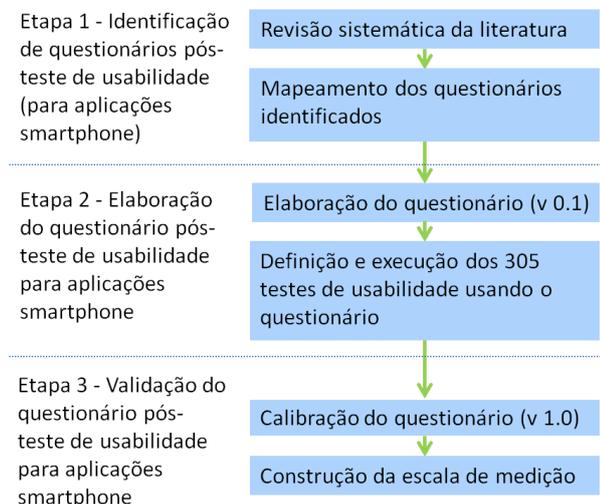


Fig. 1 - Visão geral da metodologia de pesquisa

REVISÃO SISTEMÁTICA DA LITERATURA

Com o objetivo de identificar e analisar questionários pós-teste de usabilidade existentes foi realizada uma revisão sistemática da literatura, com base em Kitchenham (2004). Para tal, foram efetuadas buscas no Google Scholar, entre os meses de abril e maio de 2013, com os termos de busca apresentados no Quadro 1.

Termo de busca
<i>User satisfaction questionnaires software</i>
<i>Questionnaires measuring satisfaction</i>
<i>User satisfaction questionnaires to measure</i>
<i>User satisfaction survey</i>
<i>Standardized questionnaire to measure satisfaction</i>
<i>User satisfaction software</i>
<i>Questionnaire to measure satisfaction interface User</i>

Quadro 1. Termos de busca usados da revisão sistemática

Foram incluídos: artigos completos, publicados em inglês, entre os anos de 1990 e 2013, com acesso livre via portal Capes, que apresentam questionários para medir a usabilidade percebida no final de testes de usabilidade. Devido ao número baixo de pesquisas voltadas para

aplicações smartphones, foram considerados também questionários para outros dispositivos, como desktop. Foram excluídos questionários para outras finalidades como, por exemplo, pós-tarefa e também aqueles que continham mais de 100 itens.

Como resultado, foram identificados 11 questionários: SUMI (KIRAKOWSKI e CORBETT, 1988), QUIS (CHIN, 1988), PUEU (DAVIS, 1989), SUS (BROOKE, 1996), PSSUQ (LEWIS, 2002), MPUQ (RYU, 2005), MoBiS-Q (MARKOVA, 2007), MoBiS-Q (VUOLLE, 2008), CSUQ (LEWIS, 2010), UMUX (FINSTAD, 2010), UMUX LITE (LEWIS, 2013). Esses questionários focam não apenas na satisfação do usuário, mas também na sua percepção sobre a eficiência, eficácia, atratividade e utilidade do sistema.

Os trabalhos identificados com foco em dispositivos móveis possuíam foco em dispositivos como PDAs, telefones celulares com teclado físico e smartphones mais simples (RYU, 2005).

Assim, mostra-se relevante o desenvolvimento de um questionário pós-teste de usabilidade específico para medir a usabilidade percebida das aplicações smartphone.

PROJETO INICIAL DO QUESTIONÁRIO PÓS-TESTE DE USABILIDADE PARA SMARTPHONES

Buscando unificar os questionários levantados, foram agrupados em uma só questão todos os itens que abordam o mesmo fator de usabilidade. Foram excluídos itens considerados mais adequados para avaliação de especialistas como: O software permite que o usuário seja econômico no número de teclas pressionadas? (KIRAKOWSKI e CORBETT, 1988). Também foram descartados itens que tratavam de fatores não relacionados à usabilidade como o desempenho do aplicativo.

Por outro lado, o conjunto foi adaptado para considerar características específicas dos smartphones, com base no MPUQ (ZHANG e ADIPAT, 2005) e Mobis-Q (AVOURIS et al, 2008). Ademais, todos os itens do SUS foram mantidos¹ propositadamente, a fim de permitir a sua comparação com o questionário proposto.

Ao final do mapeamento foi obtido um questionário com 37 itens. Devido a razões estatísticas definiu-se uma escala Likert de respostas de 4 pontos: 1-discordo, 2-discordo, 3-concordo e 4-concordo totalmente, com a opção adicional de não se aplica (N/A).

VALIDAÇÃO

Foi realizado um estudo que envolveu a aplicação do questionário em 305 testes realizados com usuários para verificar sua validade e definir uma escala de medida de usabilidade. Essa escala foi criada utilizando a Teoria de Resposta do Item (TRI). Para a criação dessa escala é necessário definir uma média e um desvio padrão, que nesse

¹ Porém, alguns de seus itens sofreram pequenas modificações para facilitar sua compreensão.

trabalho foram respectivamente 50 e 10, ou seja, utilizou-se uma escala (50,10). Posteriormente os resultados dos testes foram comparados nessa escala.

Definição e execução dos testes

Para os testes foram selecionados cinco aplicativos disponibilizados gratuitamente nas lojas das duas principais plataformas móveis atuais (GARTNER, 2013), sendo eles: Buscapé e GloboEsporte.com, para iOS e Gmail, Qual é a Resposta? e The Weather Channel para Android. A seleção foi feita de forma a obter aplicativos com diferentes níveis de usabilidade (muito alta, alta, razoável, baixa e muito baixa), de acordo com a classificação realizada por Witt (2013).

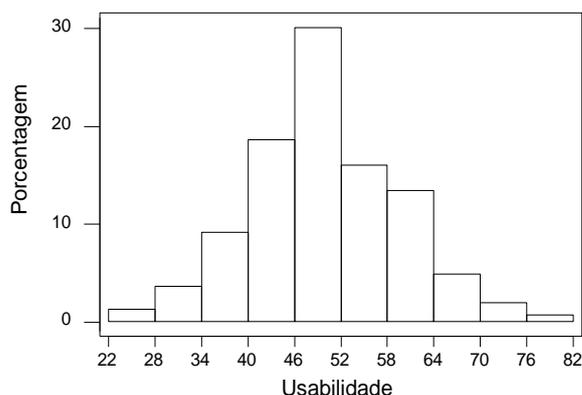
A fim de ampliar o número de participantes, os testes² foram realizados de forma remota e presencial. Os testes remotos foram realizados com a aplicação de um questionário online. Após o usuário informar a plataforma do seu smartphone, o questionário sorteava um dos cinco aplicativos predefinidos para a realização do teste. Para evitar efeito de fadiga sobre as últimas questões, a ordem dos itens de medição era alterada aleatoriamente para cada participante. Os testes presenciais foram aplicados de duas formas: a) em turmas de diferentes cursos da UFSC, onde os alunos receberam orientações e, após realizarem a tarefa em seu próprio smartphone, preencheram o questionário impresso. b) em ambiente controlado usando os smartphones³ fornecidos pelo GQS/INCoD/INE/UFSC e o questionário online.

No total foram realizados 305 testes, sendo 134 remotos, 120 em sala de aula e 51 em ambiente controlado. Todos os testes aconteceram entre setembro e outubro 2013.

Análise de Dados

Depois de organizados, os dados coletados nos testes foram analisados utilizando a Teoria de Resposta ao Item (TRI). A TRI é utilizada para medir traços latentes que não podem ser medidos diretamente, tais como usabilidade, por meio de modelos matemáticos (TEZZA et al, 2011). Esses modelos são probabilísticos e especificam uma distribuição de probabilidade sobre o conjunto de respostas possíveis aos itens com o traço latente. O modelo da TRI utilizado nesse trabalho foi o modelo logístico de 2 parâmetros em que os itens são caracterizados por um parâmetro de dificuldade e de discriminação, representando o quão os itens estão associados a usabilidade (Lord, 1980).

Os valores dos 305 testes de usabilidade foram distribuídos em uma escala crescente de usabilidade para aplicações smartphone, conforme mostra a Figura 2. Nessa escala é possível identificar, por exemplo, que 30 % dos usuários atribuíram uma nota de usabilidade entre 46 e 52 para a aplicação avaliada.



aplicativos atribuídos pelos participantes dos testes.

Considerando que um item do questionário pode ser eliminado por falta de informação suficiente para medir o fator nele investigado ou caso não esteja associado à usabilidade, foram eliminados os itens 4, 5, 20, 25, 27, 32, da versão (0.1) do questionário, apresentados no Quadro 2.

Itens eliminados
4. As mensagens de erro ajudam a corrigir os problemas.
5. Eu achei fácil usar o aplicativo com apenas uma das mãos.
20. O design de interface do aplicativo é atraente.
25. Eu precisei aprender muitas coisas para usar este aplicativo.
27. A interface é semelhante dos demais aplicativos Android/iOS.
32. Eu conseguiria realizar as tarefas com esse aplicativo em qualquer lugar. Por exemplo, em movimento andando pela rua.

Quadro 2. Itens eliminados da versão (0.1) do questionário

De acordo com a análise da TRI, todos os itens remanescentes medem o mesmo traço latente, ou seja, usabilidade.

RESULTADOS

A partir dessas alterações foi gerada a versão (1.0) do questionário, com 31 itens, chamada SURE (Smartphone Usability questionnaiRE), apresentado no Quadro 3.

Itens
1. Eu achei fácil inserir dados nestes aplicativos. Por exemplo, utilizando código QR, listas de opções etc.
2. Quando eu cometo um erro é fácil de corrigi-lo.
3. Eu achei que a ajuda/dica dada pelo aplicativo foi útil.
4. Foi fácil encontrar as informações que precisei.
5. Eu me senti no comando usando este aplicativo.
6. Eu achei adequado o tempo que levei para completar as tarefas.
7. Foi fácil de aprender a usar este aplicativo.
8. A sequência das ações no aplicativo corresponde à maneira como eu normalmente as executo. Por exemplo, a ordem de botões, campos de dados, etc.

² Esse estudo foi aprovado pelo Comitê de Ética da Universidade Federal de Santa Catarina com o número do certificado 19988413.6.0000.0121.

³ Um Samsung Galaxy Ace Duos, um TV Samsung Galaxy Y e um iPhone 4S.

9. É fácil fazer o que eu quero usando este aplicativo.
10. Foi fácil navegar nos menus e telas do aplicativo.
11. O aplicativo atende às minhas necessidades.
12. Eu recomendaria este aplicativo para outras pessoas.
13. Mesmo com pressa eu conseguiria executar as tarefas nesse aplicativo.
14. Eu achei o aplicativo consistente. Por exemplo, todas as funções podem ser realizadas de uma maneira semelhante.
15. É fácil lembrar como fazer as coisas neste aplicativo.
16. Eu usaria este aplicativo com frequência.
17. A organização dos menus e comandos de ação (como botões e links) é lógica, permitindo encontrá-los facilmente na tela.
18. Eu consegui completar as tarefas com sucesso usando este aplicativo.
19. Eu gostei de usar este aplicativo.
20. O aplicativo fornece todas as informações necessárias para completar as tarefas de forma clara e compreensível.
21. Eu achei o aplicativo muito complicado de usar.
22. Os símbolos e ícones são claros e intuitivos.
23. Eu achei os textos fáceis de ler.
24. Eu achei o aplicativo desnecessariamente complexo. Precisei lembrar, pesquisar ou pensar muito para completar as tarefas.
25. A terminologia utilizada nos textos, rótulos, títulos etc. é fácil de entender.
26. Eu precisaria de apoio de uma pessoa para usar este aplicativo.
27. Eu me senti confortável usando este aplicativo.
28. O aplicativo se comportou como eu esperava.
29. Eu achei frustrante usar este aplicativo.
30. Eu achei que as várias funções do aplicativo são bem integradas.
31. Eu me senti muito confiante usando este aplicativo.

Quadro 3 – Questionário (v 1.0)

Versão reduzida

Considerando o número relativamente alto de questões, os 31 itens do questionário foram revisados cuidadosamente pelo grupo de pesquisadores do GQS. Assim foi proposta uma versão curta do questionário com 11 itens, sendo eles: 6, 7, 8, 9, 11, 12, 19, 21, 24, 29 e 34. Essa seleção foi baseada em observações durante os testes sobre a dificuldade dos participantes em compreender e responder alguns itens do questionário. Também foram eliminadas questões que receberam respostas idênticas em todos os aplicativos, por não distinguirem bem aplicativos com boa e má usabilidade.

Em seguida, foram comparadas a dispersão e a diferença absoluta entre as respostas, assim como diferenças na curva de informações entre as duas versões e em relação ao SUS. Contudo, devido à remoção de uma das questões do SUS na etapa de validação (item 25), a comparação se deu sobre apenas 9 de seus itens.

A comparação apontou uma forte correlação entre os três instrumentos. No entanto, foi possível identificar que tanto

o SUS quanto a versão curta do SURE possuem menor eficiência que a versão extensa do questionário, sobretudo para avaliar aplicativos com alto grau de usabilidade.

A Figura 2 apresenta a diferença absoluta das notas na comparação dos três instrumentos. Neste gráfico quanto mais próximo de zero os pontos estão, menor é a quantidade de informação e menor a precisão da nota.

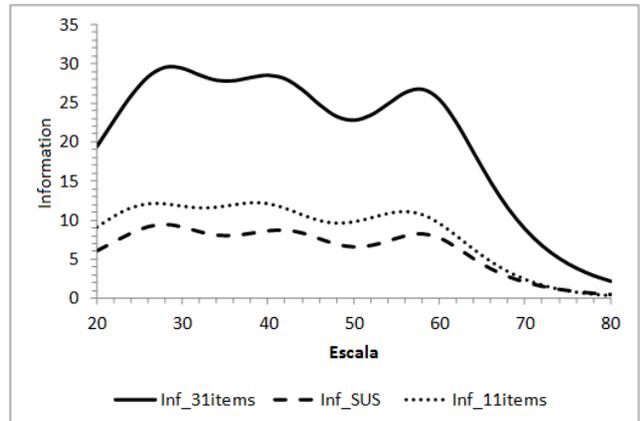


Figura 2. Curva de informação dos três instrumentos

Pode-se notar que o instrumento com 31 itens é mais informativo entre as regiões 20 e 60 da escala. Este resultado era esperado devido a maior quantidade de itens desse instrumento. Esta informação mostra que o instrumento com 31 itens é o mais preciso para gerar notas de usabilidade e diferenciar aplicativos com boa e má usabilidade.

Definição da escala de usabilidade

A partir das faixas de distribuição nas quais as notas das avaliações foram distribuídas, apresentadas na Figura 2, foi criada uma escala de usabilidade, para a qual cada nível possui uma descrição explicitando quais os fatores de usabilidade que os aplicativos classificados nessa faixa possivelmente atendam, conforme mostra o Quadro 4.

Nível	Descrição
30	Até esse nível, todos os respondentes apresentam mais possibilidade de discordar parcialmente ou totalmente dos 31 itens do instrumento.
40	Neste nível, os respondentes deixam de discordar e começam a apresentar mais possibilidade de concordar que acha fácil inserir dados no aplicativo, quando comete erro é fácil corrigir, é fácil aprender a usar o aplicativo, a sequência das ações no aplicativo corresponde à maneira como eu normalmente as executo, é fácil navegar nos menus e telas do aplicativo, acha o aplicativo consistente, é fácil lembrar como fazer as coisas neste aplicativo, consegue completar as tarefas com sucesso usando este aplicativo, não acha o aplicativo muito complicado de usar, acha os textos fáceis de ler, não acha o aplicativo desnecessariamente complexo, acha fácil de entender a terminologia utilizada nos textos, rótulos, títulos, não precisaria de

	apoio de uma pessoa para usar o aplicativo, não acha frustrante usar o aplicativo.
50	Neste nível, além de concordarem com os itens dos níveis anteriores, os respondentes também começam a apresentar mais possibilidade de concordar que acha que ajuda/dica dada pelo aplicativo foi útil, foi fácil encontrar a informação que precisava, se sentiu no comando usando o aplicativo, acha adequado o tempo que levou para completar a tarefa, é fácil fazer o que quer usando o aplicativo, o aplicativo atende às suas necessidades, recomendaria este aplicativo para outra pessoa, mesmo com pressa conseguiu executar a tarefa, usaria o aplicativo com frequência, a organização do menu e comandos da ação é lógico, gostou de usar o aplicativo, o aplicativo fornece todas as informações necessárias para completar as tarefas, os símbolos e ícones são claros e intuitivos, se sentiu confortável usando o aplicativo, o aplicativo se comportou como se esperava, as várias funções do aplicativos são bem integradas, se sentiu confiante usando o aplicativo. A partir desse nível o respondente deixa de concordar parcialmente e passa a concordar fortemente que não precisaria de apoio de uma pessoa para usar este aplicativo.
70	Nesse nível os respondentes também começam a concordar fortemente que é fácil inserir dados neste aplicativo, por exemplo, utilizando código QR, lista de opções etc, quando cometeu um erro foi fácil de corrigi-lo, a sequência das ações no aplicativo correspondeu à maneira como ele normalmente as executa, o aplicativo atende às necessidades, recomendaria este aplicativo para outras pessoas, acha o aplicativo consistente, usaria o aplicativo com frequência, achou compreensível a organização dos menus e comandos de ação (como botões e links), foi lógico, permitindo encontrá-lo facilmente na tela, gostou de usar o aplicativo, o aplicativo forneceu todas as informações necessárias para completar as tarefas, os símbolos e ícones foram claros e intuitivos, o texto foi fácil de ler, a terminologia utilizada nos textos, rótulos, títulos etc. foi fácil de entender, se sentiu confortável usando o aplicativo, o aplicativo se comportou como esperado, as várias funções do aplicativo foram bem integradas, se sentiu muito confiante usando este aplicativo.
80	É só a partir desse nível que os respondentes começam a concordar totalmente que a ajuda/dica dada pelo aplicativo foi útil. Além de concordarem totalmente com os demais itens.

Quadro 4 – Escala de usabilidade com descrição

DISCUSSÃO

A análise estatística dos dados da aplicação da versão (0.1) do questionário indica que a maioria dos itens "tradicionais"

de medição, definidos originalmente para computadores desktop, também se aplicam para medir a usabilidade de aplicações smartphone. Por outro lado, identifica inicialmente que os 6 itens eliminados não são adequados para medir a usabilidade desse tipo de aplicação, por não distinguirem aquelas com boa usabilidade das com má usabilidade.

Com base na análise, verificou-se que a versão extensa do questionário (v1.0) mostrou ser mais eficiente para a avaliação de aplicações com usabilidade entre os níveis 20 e 60, conforme mostrado na Fig.2, assim como a versão reduzida (com 11 itens). Contudo a segunda teve um desempenho menor em termos de precisão da avaliação. Ainda assim, a versão reduzida é uma alternativa para avaliações em que é inviável aplicar a versão extensa.

Da mesma forma, a proximidade entre os resultados das duas versões do questionário SURE (extensa e curta) e do SUS, um questionário reconhecidamente válido para medir a usabilidade de um aplicativo é um indício da validade dos mesmos.

CONCLUSÃO

A partir da análise estatística realizada sobre os resultados da aplicação do questionário SURE nos testes com usuários, é possível verificar que, tanto a versão longa com 31 itens quanto em a versão curta com 11 itens – consistem em ferramentas válidas para medir a usabilidade de aplicações smartphone pós-teste de usabilidade. Contudo, ainda são necessárias outras formas de validação como o cruzamento dos resultados dos 305 testes com avaliações heurísticas dos mesmos aplicativos.

Esses resultados representam um primeiro passo na avaliação da usabilidade percebida pelos usuários de aplicações smartphones. Com essa pesquisa pretende-se contribuir com a tomada de decisões durante o projeto e avaliação de usabilidade de aplicações smartphone, buscando torná-los produtos mais eficientes e agradáveis de usar.

AGRADECIMENTOS

A todos os participantes do estudo de caso e ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pelo financiamento desta pesquisa.

REFERÊNCIAS

1. Accenture, 2012. MobileWeb Watch 2012. <http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Mobile-Web-Watch-Internet-Usage-Survey-2012.pdf>.
2. Adipat, B. e Zhang D. Interface Design for Mobile Applications. *Proc. Americas Conference on Information Systems 2005*, Association for Information Systems (2005), 1862-1868.
3. Avouris N. *et al.* On Measuring Usability of Mobile Applications. *Proc. International Workshop on Meaningful Measures: Valid Useful*

- User Experience Measurement 2008*, IRIT Press (2008), 38-44.
4. Brooke, J. SUS: a “quick and dirty” usability scale. In P. Jordan, W. Thomas, B., Weerdmeester, A. e McClelland, A. L. Usability Evaluation in Industry. London: Taylor and Francis. 1996.
 5. Chin, J. P. *et al.* Development of an instrument measuring user satisfaction of the human-computer interface. *Proc. Conference on Human Factors in Computing Systems, University of Maryland, 1988*, College Park, USA (1988), 213-218.
 6. Davis, F. D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, vol. 13, no 3 (1989), 319-340.
 7. Finstad, K. The usability metric for user experience. *Interacting with Computers*, vol. 22 (2010), 323-327.
 8. Gartner. Gartner Says Smartphone Sales Accounted for 55 Percent of Overall Mobile Phone Sales in Third Quarter of 2013. <http://www.gartner.com/newsroom/id/2623415>.
 9. Hornbæk, K. Current Practice in Measuring Usability: Challenges to Usability Studies and Research, *International Journal of Human-Computer Studies*, vol. 57, no. 12, pp. 1423-1440 (2011).
 17. Markova, M. *et al.* MoBiS-Q: a tool for evaluating the success of mobile business services. *Proc. Mobile HCI*, Singapore (2007), 238-245.
 18. NIESEN, Jakob; BUDI, Raluca. Usabilidade Móvel. Elsevier: Rio de Janeiro, 2014.
 19. Ryu, Y. S. Development of Usability Questionnaires for Electronic Mobile Products and Decision Making Methods, Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA (2005).
 20. Schwab, D. P. Construct Validity in Organization Behavior. *Research in Organization Behavior*, vol. 2, Greenwich, CT: JAI Press (1980), 3-43.
 21. Seffah, A. Usability measurement and metrics: A consolidated model, *Software Quality J.*, vol. 14, no.2, Jun (2006), 159-178.
 22. Tezza, R. *et al.* Measuring web usability using item response theory: Principles, features and opportunities, *Interacting with Computers*, vol. 23, no. 2, Mar (2011), 167-175.
 - Computer Studies, vol. 64, Issue 2, February (2006), 79-102.
 10. Ketola P. e Roykee, M. The Three Facets of Usability in Mobile Handsets. *CHI Workshop, Mobile Commun: Understanding Users, Adoption & Design*, Seattle, WA, 2001.
 11. Kirakowski J. e Corbett, M. Measuring user satisfaction. *Proc. 4th Conf. of British Computer Society, Manchester*, (1988), 329-338.
 12. Kitchenham, B. A. Procedures for Performing Systematic Reviews, Keele Univ., Keele, STS, Tech. Rep. TR/SE-0401 (2004).
 13. Lewis, J. R. Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies, *Int J. of Human-Computer Interaction*, (2002), 462- 488.
 14. Lewis, J. R. Psychometric Evaluation of the CSUQ Using Data from Five Years of Usability Studies (2010).
 15. Lewis, J.R. *et al.* UMUX-LITE: when there's no time for the SUS. *Proc. CHI*, Paris (2013), 2099-2102.
 16. Lord, F. M. Applications of item response theory to practical testing problems. Mahwah, NJ: Lawrence Erlbaum Associates Inc (1980).
 23. Treeratanapon, T. Design of the Usability Measurement Framework for Mobile Applications. *Proc. Int. Conf. on Comput. and Inform. Technology*, Bangkok (2012), 71-75.
 24. Wasserman, A. I. Software Engineering Issues for Mobile Application Development. *Proc. Workshop on Mobile Software Eng.*, Santa Clara, CA (2010), 397-400.
 25. WITT, A T. Aplicação da Técnica Estatística Teoria da Resposta ao Item para avaliar um conjunto de Heurísticas de Usabilidade para dispositivos celulares *touchscreen*. 164f. Trabalho de Conclusão de Curso (Bacharel em Sistemas de Informação) – Departamento de Informática e Estatística. Universidade Federal de Santa Catarina, Florianópolis, 2013.
 26. Zhang, D. e Adipat, B. Challenges, methodologies, and issues in the usability testing of mobile applications. *Int. J. of Human-Computer Interaction*, vol. 18, no. 3 (2005), 293-308.