

Avaliação de UX em aplicações móveis: resumo de uma investigação sobre a influência de diferentes fatores em um app de compras baseado em chatbot

Walter T. Nakamura^{1,2}, Leonardo C. Marques², David Redmiles³,
Elaine H. T. de Oliveira², Tayana Conte²

¹Departamento Acadêmico de Computação (DACOM)
Universidade Tecnológica Federal do Paraná (UTFPR) – Campo Mourão, PR – Brasil

²Instituto de Computação (IComp)
Universidade Federal do Amazonas (UFAM) – Manaus, AM – Brasil

³Department of Informatics
University of California, Irvine (UCI) – Irvine, CA – USA

waltertakashi@utfpr.edu.br, redmiles@ics.uci.edu,
{lcm,elaine,tayana}@icompu.ufam.edu.br

Abstract. *User eXperience (UX) evaluations are vital in software development to guide future releases. Thus, it is crucial to pinpoint factors altering user judgments to understand the results better. This study investigates the effect of interaction sequencing, prior experience, and the number of problems on overall satisfaction and key UX dimensions (pragmatic and hedonic) by applying three methods to assess a chatbot-driven mobile shopping app. Results unveiled that previous experience and the number of problems can affect UX evaluations, mainly on the pragmatic dimension. Interaction sequencing, in turn, did not significantly influence UX.*

Resumo. *Avaliar a Experiência do Usuário (UX) no desenvolvimento de software é vital para direcionar lançamentos futuros. Assim, é crucial identificar os fatores que alteram os julgamentos do usuário para compreender melhor os resultados. Este estudo investiga o efeito do sequenciamento da interação, da experiência anterior e do número de problemas na satisfação geral e nas principais dimensões de UX (pragmática e hedônica), aplicando três métodos para avaliar um aplicativo de compras móvel baseado em chatbot. Os resultados revelaram a influência da experiência anterior e do número de problemas na UX, principalmente na dimensão pragmática. O sequenciamento da interação, por sua vez, não influenciou significativamente a UX.*

1. Introdução

O sucesso dos produtos interativos depende da disposição dos usuários, ligada ao seu prazer percebido [Cockburn et al. 2017]. Em meio à concorrência acirrada, as empresas de desenvolvimento de software vem priorizando a experiência do usuário (UX) para desenvolver produtos e serviços de sucesso [Alves et al. 2014, Ardito et al. 2014]. Nesse contexto, as avaliações de UX são essenciais para avaliar a qualidade das aplicações. No entanto, a maneira como os usuários formulam tais avaliações permanece incerto [Kujala et al. 2017].

Pesquisas recentes sugerem que os usuários podem avaliar positivamente a experiência apesar dos problemas de usabilidade [Nakamura et al. 2019] e das emoções negativas vivenciadas [Bopp et al. 2016], indicando uma possível limitação dos métodos existentes na captura de fatores que podem estar pesando mais na avaliação de UX. Dado que decisões de design dependem de tais resultados, compreender os fatores que podem influenciar nas avaliações dos usuários torna-se fundamental.

Este artigo apresenta um estudo para investigar o efeito de três fatores (número de problemas, sequenciamento da interação e experiência anterior) nas duas principais dimensões de UX (pragmática e hedônica) e na satisfação geral ¹. Para isso, foram empregados três métodos de avaliação de UX. Os resultados indicam diferenças significativas nas avaliações de UX, influenciadas por alguns dos fatores estudados.

2. Estudo Experimental

O estudo consistiu na avaliação de um aplicativo de compras online baseado em chatbot e envolveu 33 participantes, incluindo inspetores e usuários finais. O grupo de inspeção contou com 11 participantes, majoritariamente estudantes de graduação em Ciência da Computação da Universidade Federal do Amazonas (UFAM), e o grupo de usuários finais com 22 participantes de diferentes cursos.

Foram empregados dois tipos de métodos: inspeção e teste. Para a inspeção, foi utilizado o método UX-Tips [Marques et al. 2021] que, diferentemente dos métodos de inspeção tradicionais, considera fatores de UX como estética e emoção. Para o teste, foi selecionado o método *Think-Aloud*, um dos métodos mais utilizados em testes com usuários [Alhadreti and Mayhew 2018]. Para complementar os resultados, foi utilizado a versão reduzida do método UEQ (*User Experience Questionnaire*) [Schrepp et al. 2017], composta por 8 itens avaliando as dimensões Pragmática (PQ) e Hedônica (HQ), juntamente com a escala de valência do método *Self-Assessment Manikin* (SAM) [Bradley and Lang 1994] para avaliar, respectivamente, a percepção dos participantes sobre a UX e a satisfação geral.

Para avaliar o efeito do sequenciamento da interação [Cockburn et al. 2015], foram elaborados dois scripts: um com início negativo seguido de final positivo (condição *+fim*) e outro com início positivo seguido de final negativo (condição *-fim*). Ambas as condições possuíam exatamente as mesmas tarefas, porém, em diferentes ordens. Na condição *+fim*, a última tarefa seria menos demorada e menos complexa, enquanto na condição *-fim*, a última tarefa seria mais demorada e mais complexa.

Após a realização das etapas anteriores, dois pesquisadores revisaram e consolidaram os problemas. Ao término do processo foi gerado um relatório apresentando os principais problemas identificados (aqueles com maiores índices de ocorrência) aos stakeholders, os quais reformularam a aplicação.

3. Resultados

3.1. Inspeção vs. Teste

Foram identificados 173 problemas únicos, sendo 128 (74%) detectados na inspeção, 78 (45,1%) no teste e 33 (19,1%) em ambos. Embora a inspeção não tenha coberto parte

¹Este trabalho foi originalmente publicado em: [Nakamura et al. 2022].

dos problemas identificados no teste, a técnica UX-Tips possibilitou a captura de aspectos emocionais dos inspetores. Utilizando as 27 categorias de experiências emocionais do trabalho de Cowen e Keltner (2017) como base, foi verificado se os problemas continham aspectos emocionais nos relatos ou não. No geral, a inspeção resultou em um maior número de problemas relatados utilizando termos emocionais. Foram identificados 28 problemas relatados por 10 de 11 inspetores (91%) que descreveram como o problema afetou a experiência. Em relação ao teste, foram identificadas 16 descrições de problemas de 12 dentre 22 usuários (55%) que apresentavam termos emocionais. Embora os participantes do teste tenham sido encorajados a expressarem seus pensamentos e suas emoções durante o teste, muitos não o fizeram.

3.2. Avaliação de UX da Aplicação

Para avaliar a UX, calculou-se a média de cada dimensão avaliada pela UEQ, bem como a média da escala de Valência do SAM para avaliar a satisfação geral. As próximas seções apresentam os resultados de acordo com os fatores investigados.

3.2.1. Método de Avaliação vs. Avaliação de UX

Os resultados foram contrastantes entre inspetores e usuários. Enquanto os inspetores tenderam a uma avaliação mais negativa, os usuários foram mais positivos sobre a sua experiência com a aplicação, resultando em diferenças na satisfação e na dimensão PQ, sendo esta última significativa. Por sua vez, a diferença na dimensão HQ não foi significativa, com ambos os grupos reconhecendo o aspecto inovador da aplicação. Considerando que a inspeção foca na identificação de problemas de interface e interação, os problemas identificados pelos inspetores durante a avaliação da aplicação pode ter influenciado a percepção deles sobre sua qualidade pragmática.

Para avaliar o efeito do método na avaliação de UX dos inspetores, foi conduzido uma regressão linear para investigar a relação entre as dimensões de UX, satisfação geral e número de problemas. A regressão geral para a satisfação foi estatisticamente significativa, explicando 69,4% da variância. A PQ previu significativamente a satisfação geral, mas não a HQ. Estes resultados indicam que a satisfação dos inspetores foi influenciada pela percepção deles sobre a PQ. Em relação ao número de problemas, a regressão geral não foi estatisticamente significativa. Assim, não há uma relação direta entre a quantidade de problemas e a avaliação de UX, ou seja, uma maior quantidade de problemas identificados pelo inspetor não necessariamente levará a uma avaliação de UX mais negativa.

3.2.2. Análise por Sequenciamento da Interação

Para essa análise, os dados foram analisados de acordo com o grupo (inspeção e teste) e a condição (*+fim* e *-fim*). Em relação ao teste, os resultados foram bem similares, com pequenas diferenças entre as duas condições. Participantes na condição *-fim* deram notas um pouco mais baixas, principalmente na dimensão HQ. Contudo, a diferença não foi significativa nem para as dimensões PQ e HQ, nem para a satisfação geral.

Em relação ao grupo de inspeção, as diferenças foram maiores, embora não significativas. Inspetores na condição *-fim* tenderam a fornecer avaliações mais negativas. A

tarefa mais longa ao final com muitas etapas a serem realizadas e inspecionadas pode ter afetado a percepção dos inspetores sobre a aplicação, levando a avaliações mais negativas devido ao efeito de recência, que ocorre quando a última experiência é mais influente no julgamento da pessoa [Hands and Avons 2001].

3.2.3. Análise por Experiência Anterior

Os resultados não apontaram diferença significativa entre os grupos na percepção sobre os aspectos inovadores da aplicação (HQ). Contudo, a experiência anterior afetou a percepção sobre a qualidade pragmática (PQ) e satisfação geral. Ao contrário do esperado, usuários experientes deram notas mais baixas para a dimensão PQ e satisfação, enquanto aqueles sem experiência anterior deram notas mais altas nessas dimensões. O número de problemas enfrentados durante a interação foi semelhante para ambos os grupos, sugerindo que a experiência com aplicações similares não resultou em uma maior capacidade de utilizar a aplicação, cuja forma de interação difere das aplicações de compras convencionais. Os resultados da avaliação da aplicação similar revelaram que, em geral, as lembranças dos participantes sobre a interação anterior foi positiva, principalmente na dimensão PQ. Por outro lado, é interessante mencionar que os participantes consideraram o aplicativo avaliado como mais inventivo do que os utilizados anteriormente.

Estes resultados podem ser explicados pela Teoria da Confirmação de Expectativas [Hossain and Quaddus 2012, Zeithaml et al. 1990], em que os usuários formam expectativas iniciais com base na experiência anterior e sua satisfação é influenciada pelo fato do produto atendê-las ou excedê-las. Usuários experientes tinham grandes expectativas com base em sua experiência anterior, tornando mais difícil para a nova aplicação atender a essas expectativas e resultando em menor satisfação. Usuários inexperientes tiveram percepções variadas e nenhuma base para comparação, levando a uma relação menos consistente entre a satisfação e as dimensões de UX. Assim, mudanças drásticas na forma como o usuário costuma interagir devem ser evitadas, pois podem levar a percepções negativas principalmente quando acompanhadas de dificuldades no seu uso.

4. Conclusões

Este artigo apresentou um estudo para investigar o impacto da experiência anterior, sequenciamento da interação e número de problemas nas percepções do usuário sobre a UX. Para isso, foram adotados métodos de inspeção, teste e um questionário de UX.

Os resultados apontaram que as percepções sobre a UX podem ser afetadas por diferentes fatores, particularmente o método de avaliação e a experiência anterior. Assim, é importante que pesquisadores e profissionais estejam cientes de seus impactos durante o processo de desenvolvimento e avaliação do software.

A diferença entre as avaliações de UX de usuários com e sem experiência anterior abre oportunidades para a realização de mais estudos para investigar melhor o efeito desse fator. Pesquisas futuras podem investigar o seu efeito sob uma perspectiva longitudinal, de forma a investigar se e como as percepções iniciais mudam ao longo do tempo entre os dois grupos após se familiarizarem com a nova forma de interação. Com relação ao sequenciamento da interação, pode-se investigá-lo em outros tipos de aplicativos para compreender melhor o seu impacto na UX.

Referências

- Alhadreti, O. and Mayhew, P. (2018). Rethinking thinking aloud: A comparison of three think-aloud protocols. pages 1–12, Montreal QC, Canada. the 2018 CHI Conference, ACM Press.
- Alves, R., Valente, P., and Nunes, N. J. (2014). The state of user experience evaluation practice. pages 93–102, Helsinki, Finland. the 8th Nordic Conference, ACM Press.
- Ardito, C., Buono, P., Caivano, D., Costabile, M. F., and Lanzilotti, R. (2014). Investigating and promoting ux practice in industry: An experimental study. *International Journal of Human-Computer Studies*, 72(6):542–551.
- Bopp, J. A., Mekler, E. D., and Opwis, K. (2016). Negative emotion, positive experience?: Emotionally moving moments in digital games. pages 2996–3006, San Jose California USA. CHI'16: CHI Conference on Human Factors in Computing Systems, ACM.
- Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59.
- Cockburn, A., Quinn, P., and Gutwin, C. (2015). Examining the peak-end effects of subjective experience. pages 357–366, Seoul, Republic of Korea. the 33rd Annual ACM Conference, ACM Press.
- Cockburn, A., Quinn, P., and Gutwin, C. (2017). The effects of interaction sequencing on user experience and preference. *International Journal of Human-Computer Studies*, 108:89–104.
- Cowen, A. S. and Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, 114(38):E7900–E7909.
- Hands, D. S. and Avons, S. E. (2001). Recency and duration neglect in subjective assessment of television picture quality. *Applied Cognitive Psychology*, 15(6):639–657.
- Hossain, M. A. and Quaddus, M. (2012). Expectation–confirmation theory in information system research: A review and analysis. *Information systems theory*, pages 441–469.
- Kujala, S., Mugge, R., and Miron-Shatz, T. (2017). The role of expectations in service evaluation: A longitudinal study of a proximity mobile payment service. *International Journal of Human-Computer Studies*, 98:51–61.
- Marques, L., Matsubara, P. G., Nakamura, W. T., Ferreira, B. M., Wiese, I. S., Gadelha, B. F., Zaina, L. M., Redmiles, D., and Conte, T. U. (2021). Understanding ux better: A new technique to go beyond emotion assessment. *Sensors*, 21(21).
- Nakamura, W. T., de Oliveira, E. H. T., and Conte, T. (2019). Negative emotions, positive experience: What are we doing wrong when evaluating the ux? pages 1–6, Glasgow, Scotland Uk. Extended Abstracts of the 2019 CHI Conference, ACM Press.
- Nakamura, W. T., Marques, L. C., Redmiles, D., de Oliveira, E. H., and Conte, T. (2022). Investigating the influence of different factors on the ux evaluation of a mobile application. *International Journal of Human–Computer Interaction*, pages 1–21.

- Schrepp, M., Hinderks, A., and Thomaschewski, J. (2017). Design and evaluation of a short version of the user experience questionnaire (ueq-s). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(6):103.
- Zeithaml, V. A., Parasuraman, A., Berry, L. L., and Berry, L. L. (1990). *Delivering quality service: Balancing customer perceptions and expectations*. Simon and Schuster.