

Proposta de questionário padronizado para avaliação de Experiência de Usuário em Realidade Virtual imersiva

Jean F. P. Cheiran^{1,2}, Denise R. Bandeira³, Marcelo S. Pimenta²

¹ Universidade Federal do Pampa (UNIPAMPA) – campus Alegrete
Alegrete – RS – Brasil

²Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brazil

³Instituto de Psicologia – Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brasil

jeancheiran@unipampa.edu.br, bandeira@ufrgs.br, mpimenta@inf.ufrgs.br

Abstract. *As Virtual Reality (VR) applications get popular, the need for quick, inexpensive, and reliable User Experience (UX) evaluation methods also increases. However, the lack of standardized questionnaires for assessing the UX in VR leads researchers and practitioners to use less accurate evaluation strategies. Hence, this study aims to develop a standardized questionnaire in Brazilian Portuguese for assessing the UX in immersive VR and to evaluate the questionnaire's psychometric properties. Therefore, we establish a theoretical model, create a questionnaire based on content analysis and experts' feedback, and evaluate the suitability of the measurement model (the questionnaire) and the structural model (the theoretical model).*

Resumo. *Conforme os aplicativos em Realidade Virtual (RV) se popularizam, aumenta a necessidade de avaliação da Experiência do Usuário (UX) de forma rápida, barata e confiável. Contudo, a falta de questionários padronizados para avaliar UX em RV leva pesquisadores e profissionais a utilizarem estratégias de avaliação menos precisas. Assim, esse trabalho tem como objetivo desenvolver um questionário padronizado em português brasileiro para avaliação de UX em RV imersiva com propriedades psicométricas avaliadas. Para isso, estabelecemos um modelo teórico, criamos um questionário a partir de análise de conteúdo e feedback de especialistas, e avaliamos a adequação do modelo de medição (questionário) e do modelo estrutural (modelo teórico).*

1. Introdução

A Realidade Virtual (RV), enquanto “ambiente artificial que é experienciado através de estímulos sensoriais fornecidos por um computador e no qual as ações de alguém determinam parcialmente o que acontece”¹, tem passado por um renascimento nos últimos anos. A comercialização de dispositivos imersivos de preço acessível, os amplos investimentos em metaverso e a intensificação de teletrabalho estão entre os fatores que impulsionam a criação de novos ambientes de RV para pesquisa, lazer, trabalho etc.

¹<https://www.merriam-webster.com/dictionary/virtual%20reality>

Tendo em vista que uma parte fundamental do desenvolvimento de sistemas interativos é a avaliação de Interação Humano-Computador (IHC) [Barbosa et al. 2021], a RV exige a consideração de aspectos particulares que extrapolam a qualidade em uso como, por exemplo, a sensação de presença² e *VR sickness*³. Essa perspectiva mais ampla nos leva à noção de Experiência de Usuário (UX, do inglês *User eXperience*) que engloba todas as percepções e respostas do usuário que resultam do uso ou da expectativa de uso de um produto [ISO 2019].

Quantitativamente, a UX pode ser estimada por meio de questionários e medidas psicofisiológicas [Vermeeren et al. 2010, Rivero and Conte 2017]. Embora o debate sobre abordagens quantitativas e qualitativas seja intenso [Law 2011] e sua escolha dependa dos objetivos e do contexto da avaliação [Barbosa et al. 2021], questionários são amplamente usados [Law 2011, Rivero and Conte 2017]. Nesse sentido, fatores que favorecem o uso de questionários são a maior dificuldade de análise de medidas qualitativas [Rivero and Conte 2017], a necessidade de treinamento e experiência por parte de entrevistadores e observadores [Barbosa et al. 2021], a imprecisão de medidas psicofisiológicas isoladas [Vermeeren et al. 2010] e a possibilidade de avaliação da qualidade psicométrica dos questionários do ponto de vista de validade, fidedignidade e sensibilidade [Sauro and Lewis 2016]. Esse último fator é particularmente determinante na criação de questionários padronizados, ou seja, instrumentos projetados para uso recorrente, com itens em ordem e formato específicos, incluindo instruções para aplicação e interpretação de resultados [Sauro and Lewis 2016].

Embora questionários sejam uma ferramenta adequada para muitos cenários de avaliação de UX, é possível identificar três problemas para profissionais e pesquisadores brasileiros que queiram utilizá-los na avaliação de aplicações em RV: (a) a menor precisão nos resultados obtidos por meio de questionários *ad hoc* [Hornbaek and Law 2007], (b) a falta de questionários padronizados de avaliação traduzidos para português brasileiro e adequadamente validados [Melo and Darin 2019] e (c) as fragilidades dos questionários padronizados existentes para avaliar UX no contexto de RV de uma forma holística (tema discutido na seção 2).

Dessa forma, o objetivo desse trabalho é desenvolver um questionário padronizado em português brasileiro para avaliação de Experiência de Usuário em Realidade Virtual imersiva com base na Teoria Clássica dos Testes.

2. Trabalhos Relacionados

Diversos questionários padronizados utilizados nos processos de avaliação já foram mapeados na literatura científica de IHC e RV. Dentre os esforços para identificação de métodos e ferramentas para avaliação de UX em um contexto geral destacam-se mapeamentos e revisões sistemáticas da área [Vermeeren et al. 2010, Rivero and Conte 2017, Pettersson et al. 2018]. No contexto de avaliação de UX especificamente para ambientes de RV, destacam-se revisões de UX em uma perspectiva holística [Kim et al. 2020] e revisões de componentes específicos de RV [Chang et al. 2020,

²Presença é a sensação de estar em um lugar diferente daquele em que o corpo físico está, sendo uma reação psicológica e subjetiva a ambientes imersivos [Grassini and Laumann 2020]

³*VR sickness* representa um conjunto de sintomas desagradáveis similares a *motion sickness* e decorrentes do uso de Realidade Virtual, incluindo desorientação e enjoo [Chang et al. 2020].

Grassini and Laumann 2020].

Considerando que a UX é composta por inúmeros aspectos determinados pelo contexto [Schrepp and Thomaschewski 2019], o uso de questionários padronizados de UX genérica para avaliar experiências em RV frequentemente é insuficiente. Questionários como, por exemplo, o AttrakDiff [Hassenzahl et al. 2003] e o User Experience Questionnaire (UEQ) [Schrepp et al. 2017] são bastante utilizados, contudo a interpretação de resultados que os utilizem é limitada quando consideramos que não abordam explicitamente fatores importantes da experiência em RV. Nesses casos, assume-se que a influência de componentes específicos de RV nos resultados é irrelevante ou que os componentes específicos de RV (não medidos) têm correlação significativa com outros componentes (medidos). Tendo em vista que não há consenso sobre as relações entre componentes de UX em RV [Cheng et al. 2014, Tcha-Tokey et al. 2018] e que a UX em diferentes tipos de software dependem de aspectos distintos [Schrepp and Thomaschewski 2019], qualquer um dos pressupostos pode gerar erros na interpretação dos resultados.

Outra abordagem comum na avaliação de UX em RV é o uso de múltiplos questionários padronizados no processo de avaliação, embora isso possa levar a outros problemas causados por formatos de item diferentes e incertezas sobre como um escore total de UX poderia ser calculado a partir de escores parciais [Schrepp and Thomaschewski 2019].

Poucos questionários padronizados para avaliação de UX em RV foram identificados ao longo do desenvolvimento desse trabalho, sendo os mais abrangentes o Virtual Experience Test (VET) [Chertoff et al. 2010], o User Experience in Immersive Virtual Environments questionnaire (UXIVE) [Tcha-Tokey et al. 2016] e um questionário sem nome específico [Cheng et al. 2014]. Contudo, há algumas fragilidades nos processos de construção e validação desses instrumentos que incluem a falta de evidências de validade no questionário UXIVE, a falta de estimativas de fidedignidade no VET, o uso de coeficientes de fidedignidade controversos no UXIVE, a investigação de dimensionalidade por meio de técnicas pouco adequadas (Análise de Componentes Principais) no VET e no questionário de Cheng et al., e a ausência de instruções para calcular os escores no UXIVE e no questionário de Cheng et al.

Há ainda propostas de questionários padronizados para avaliar UX em contextos específicos (e.g., reabilitação, técnicas de locomoção etc.), contudo eles não serão discutidos por serem desenvolvidos a partir de modelos particulares de UX e por incluírem componentes que não se aplicam à maioria dos cenários de RV.

Finalmente, cabe destacar que a quantidade de instrumentos de avaliação de UX ou de seus componentes em português é bastante pequena [Melo and Darin 2019], tendo em vista que processos sistemáticos de tradução e adaptação precisam ser acompanhados de novas evidências de validade e estimativas de fidedignidade que demonstrem que as propriedades do instrumento não foram modificadas [AERA et al. 2014].

3. Materiais e Métodos

A Figura 1 apresenta uma visão geral desse estudo, dividida em atividades executadas em duas fases: antes e depois da defesa da proposta de tese. Essa pesquisa se caracteriza como um estudo correlacional, quantitativo e experimental [Sampieri et al. 2013].

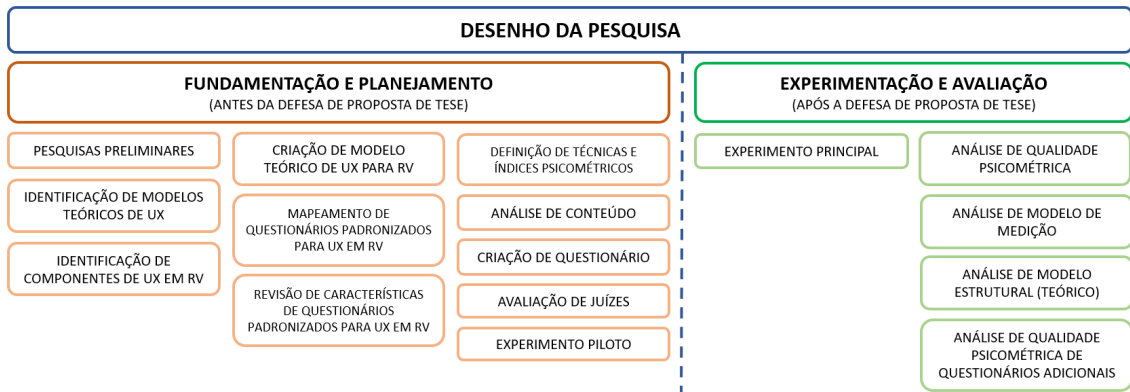


Figura 1. Desenho da pesquisa com as principais atividades.

As pesquisas preliminares, a identificação de modelos teóricos de UX e a identificação de componentes de UX em RV ocorreram de forma exploratória a partir de artigos de referência [Blythe et al. 2007, Chertoff et al. 2010, Cheng et al. 2014, Tcha-Tokey et al. 2016, Tcha-Tokey et al. 2018, Minge et al. 2017] e revisões sobre o tema [Vermeeren et al. 2010, Rivero and Conte 2017, Pettersson et al. 2018, Kim et al. 2020, Chang et al. 2020, Grassini and Laumann 2020].

O modelo teórico proposto é uma adaptação do modelo Components of User Experience (CUE) para UX [Mahlke 2008]. A Figura 2 apresenta o modelo (ainda não confirmado) que introduz elementos de *VR sickness* e presença. Mudanças adicionais incluem a reestruturação de subfatores e a simplificação de dimensões. As adaptações foram fundamentadas em modelos de UX para RV e estudos sobre relações entre componentes da experiência em RV.

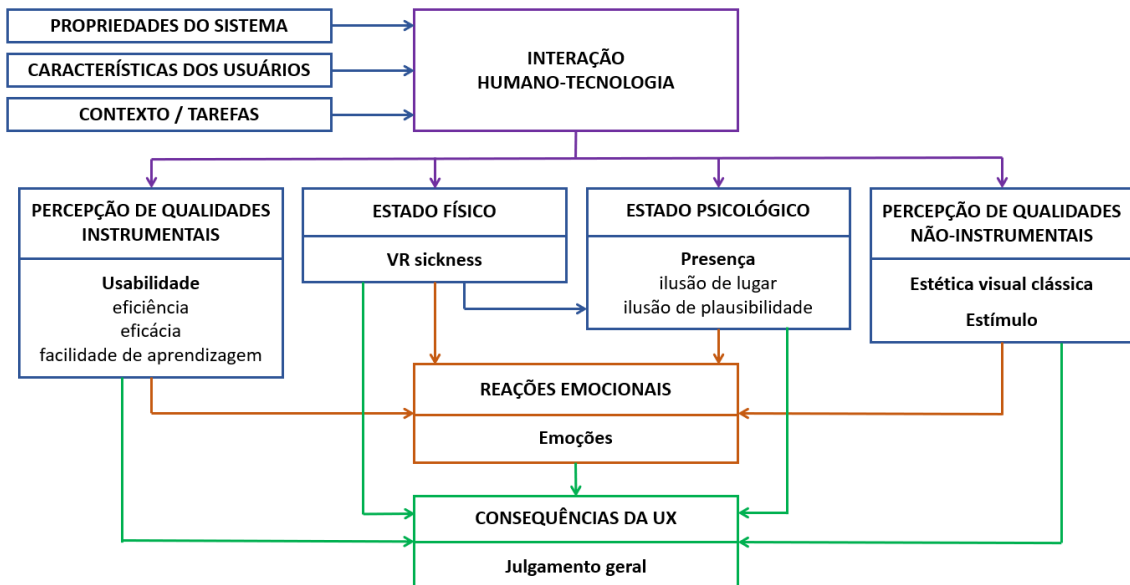


Figura 2. Modelo teórico de UX para RV adaptado de [Mahlke 2008].

Um mapeamento sistemático de questionários validados para avaliação de UX em ambientes de RV envolveu cinco pesquisadores e a revisão das características desses

questionários (dimensões, formato dos itens, evidências de validade e estimativas de fidedignidade, ambiente de RV nos experimentos etc.) guiou a investigação e as decisões metodológicas para a criação de nosso questionário.

Decisões importantes sobre a abordagem psicométrica incluem: (a) fundamentação clara do questionário na Teoria Clássica dos Testes (TCT) [AERA et al. 2014, Hutz et al. 2015], (b) investigação de validade em uma perspectiva de diferentes fontes de evidência [AERA et al. 2014, Hutz et al. 2015], (c) adoção de estimativas de fidedignidade para um modelo estimado como multidimensional e congênico [Cho 2016] e (d) estimativa do tamanho da amostra que considere aspectos dos modelos de medição e estrutural [Knekta et al. 2018].

A criação do questionário parte de uma análise de conteúdo [Moraes 1999] dos principais instrumentos utilizados para avaliar UX (ou seus sub componentes) em RV. A escolha dos instrumentos se baseou no mapeamento realizado e nas revisões de outros autores. Além de serem incluídos os questionários que avaliam UX em uma perspectiva holística em RV, a análise de conteúdo contém instrumentos de avaliação populares e amplamente usados em experimentos de RV nas perspectivas de presença, *VR sickness*, usabilidade, emoção/afeto, estética visual, carga física e cognitiva, aceitação e uso de tecnologia, motivação e engajamento. A redação dos itens preliminares segue as recomendações de Hutz, Bandeira e Trentini [Hutz et al. 2015].

O questionário criado é composto por itens no estilo Likert: uma faixa de resposta simétrica, etiquetada e numerada na qual o usuário indica sua concordância com uma afirmação. A faixa de resposta contém 7 pontos para viabilizar etiquetas coerentes e não ambíguas, sendo também equilibrada em relação à facilidade de uso e à expressividade e possuindo características psicométricas levemente melhores que faixas com menos pontos [Preston and Colman 2000].

A avaliação de juízes tem objetivo de identificar problemas de qualidade e pertinência dos itens para medir aquela dimensão [Hutz et al. 2015]. A análise qualitativa das respostas, em vez de apenas um cálculo de coeficiente de concordância, permite o refinamento de itens com base nos comentários dos juízes.

Um experimento piloto é usado no lugar de grupos focais [Hutz et al. 2015], uma vez que entender as questões do questionário depende do uso de RV (i.e., alguém que nunca experimentou RV pode não entender efetivamente o que significa ‘conseguir interagir com as coisas como se fossem reais’). Esses participantes, depois de utilizar aplicativos diferentes em RV imersiva por meio de um óculos de RV e responder o questionário após cada uso, são inquiridos sobre a clareza de cada item. O *feedback* desses participantes permite ajustes mais finos no questionário. Os aplicativos incluem um ambiente tutorial lúdico, um jogo de ritmo e um aplicativo de pintura.

O experimento principal é praticamente idêntico ao experimento piloto, embora nenhum *feedback* sobre a qualidade do questionário seja coletado. Além disso, questionários adicionais [de Carvalho et al. 2011, Martins et al. 2015, Silva et al. 2016] são respondidos por cada participante. Os questionários são impressos, pois são mais flexíveis para replicação dos formatos originais. A técnica *latin squares* é usada para que os participantes utilizem aplicativos e respondam questionários em diferentes ordens.

A análise da qualidade psicométrica se concentra em estimativas de fidedignidade

congenérica (coeficientes ômega hierárquico e total) e evidências de validade baseadas no conteúdo do teste (argumentação sobre itens baseada em modelos teóricos e questionários existentes, avaliação por juízes, avaliação em estudo piloto), na estrutura interna (análises fatoriais, correlação intraclasse, correlação item-total, análise de invariância da medida) e em relações com outras variáveis (correlação entre dimensões do questionário e outros questionários existentes).

A análise do modelo de medição é feita através de Análise Fatorial Confirmatória (AFC) [Knekta et al. 2018] e a comparação de modelos com diferentes organizações (multifator, fator de segunda ordem e bifator) é realizada por meio de índices de adequação absolutos e incrementais. No caso da obtenção de índices de baixa qualidade em todos os modelos e sem uma explicação clara, uma Análise Fatorial Exploratória (AFE) [Knekta et al. 2018] é utilizada para identificar o modelo de medição mais adequado para os dados. A quantidade de fatores no modelo é determinada pela análise do gráfico de escarpa (*scree plot*) dos *eigenvalues* e pela análise paralela de Horn.

Dois análises de invariância da medida são consideradas: uma relacionada aos aplicativos usados e outra relacionada à experiência dos participantes com ambientes de RV. Finalmente, a investigação sobre o modelo estrutural é realizada através de análise de caminhos (*path analysis*).

Um protocolo de biossegurança relacionado à Covid-19 foi adotado, assim como um processo de triagem que visa identificar condições que possam colocar os participantes em risco pela utilização de RV. Esse estudo foi autorizado pelo Comitê de Ética em Pesquisa da Universidade Federal do Rio Grande do Sul por meio da Plataforma Brasil sob número 53245721.1.0000.5347.

4. Considerações Preliminares

Atualmente, essa pesquisa encontra-se na fase de Fundamentação e Planejamento (Figura 1) e as atividades concluídas são: pesquisas preliminares, identificação de modelos e componentes de UX, criação de modelo teórico, mapeamento de questionários e revisão de suas características, definição de técnicas psicométricas, análise de conteúdo, criação do questionário e avaliação de juízes.

A principal contribuição desse trabalho para desenvolvedores de aplicações em RV no Brasil é o questionário padronizado que permite uma estimação da experiência subjetiva dos usuários dessas aplicações. Adicionalmente, espera-se oferecer para pesquisadores brasileiros um modelo teórico de UX em RV que viabilize novas investigações (incluindo a extensão do questionário desenvolvido) e, também, descrever um processo detalhado de criação e avaliação psicométrica de um questionário padronizado.

5. Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referências

AERA, APA, and NCME (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association, Washington, DC, EUA.

- Barbosa, S. D. J., Silva, B. S. d., Silveira, M. S., Gasparini, I., Darin, T., and Barbosa, G. D. J. (2021). *Interação Humano-Computador e Experiência do Usuário*. autopublicação, s.l.
- Blythe, M., Hassenzahl, M., Law, E., and Vermeeren, A. (2007). An analysis framework for user experience (ux) studies: a green paper. In *COST294-MAUSE affiliated workshop*, pages 1–5, Lancaster, RU. COST294-MAUSE.
- Chang, E., Kim, H. T., and Yoo, B. (2020). Virtual Reality Sickness: A Review of Causes and Measurements. *International Journal of Human-Computer Interaction*, 36(17):1658–1682.
- Cheng, L. K., Chieng, M. H., and Chieng, W. H. (2014). Measuring virtual experience in a three-dimensional virtual reality interactive simulator environment: A structural equation modeling approach. *Virtual Reality*, 18(3):173–188.
- Chertoff, D. B., Goldiez, B., and LaViola, J. J. (2010). Virtual experience test: A virtual environment evaluation questionnaire. In *IEEE Virtual Reality Conference*, Waltham, MA, EUA. IEEE.
- Cho, E. (2016). Making Reliability Reliable: A Systematic Approach to Reliability Coefficients. *Organizational Research Methods*, 19(4):651–682.
- de Carvalho, M. R., da Costa, R. T., and Nardi, A. E. (2011). Simulator Sickness Questionnaire: tradução e adaptação transcultural. *Jornal Brasileiro de Psiquiatria*, 60(4):0–5.
- Grassini, S. and Laumann, K. (2020). Questionnaire measures and physiological correlates of presence: A systematic review. *Frontiers in Psychology*, 11:349.
- Hassenzahl, M., Burmester, M., and Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & Computer 2003*, pages 187–196. Vieweg+Teubner Verlag, Wiesbaden, Alemanha.
- Hornbaek, K. and Law, E. L.-C. (2007). Meta-analysis of correlations among usability measures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 617–626, New York, NY, EUA. ACM.
- Hutz, C. S., Bandeira, D. R., and Trentini, C. M., editors (2015). *Psicometria*. Artmed, Porto Alegre, RS, Brasil.
- ISO (2019). ISO 9241-210:2019 - ergonomics of human-system interaction - part 210: Human-centred design for interactive systems. Standard, International Organization for Standardization, Genebra, Suíça.
- Kim, Y. M., Rhiu, I., and Yun, M. H. (2020). A Systematic Review of a Virtual Reality System from the Perspective of User Experience. *International Journal of Human-Computer Interaction*, 36(10):893–910.
- Knekta, E., Runyon, C., and Eddy, S. (2018). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE—Life Sciences Education*, 18(1):1–17.
- Law, E. L. (2011). The measurability and predictability of user experience. In *Proceedings of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, page 1–10, New York, NY, EUA. ACM.

- Mahlke, S. (2008). *User Experience of Interaction with Technical Systems*. Doktor der ingenieurwissenschaften, Technische Universität Berlin, Fakultät V, Alemanha.
- Martins, A. I., Rosa, A. F., Queirós, A., Silva, A., and Rocha, N. P. (2015). European Portuguese validation of the System Usability Scale (SUS). *Procedia - Procedia Computer Science*, 67:293–300.
- Melo, B. and Darin, T. (2019). Scope and definition of user experience in brazil: A survey to explore community's perspectives. In *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems*, pages 1–11, New York, NY, EUA. ACM.
- Minge, M., Thüring, M., Wagner, I., and Kuhr, C. V. (2017). The mecue questionnaire: A modular tool for measuring user experience. In *Advances in Ergonomics Modeling, Usability & Special Populations*, pages 115–128, Cham, Suíça. Springer.
- Moraes, R. (1999). Análise de conteúdo. *Revista Educação*, 22(37):7–32.
- Pettersson, I., Lachner, F., Frison, A.-K., Riener, A., and Butz, A. (2018). A Bermuda Triangle? - A Review of Method Application and Triangulation in User Experience Evaluation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–16, Montréal, QC, Canadá. ACM.
- Preston, C. C. and Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1):1–15.
- Rivero, L. and Conte, T. (2017). A systematic mapping study on research contributions on ux evaluation technologies. In *Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems*, pages 1–10, New York, NY, EUA. ACM.
- Sampieri, R. H., Collado, C. F., and Lucio, M. P. B. (2013). *Metodologia de pesquisa*. Grupo A, Porto Alegre, RS, Brasil.
- Sauro, J. and Lewis, J. (2016). *Quantifying the User Experience: Practical Statistics for User Research*. Elsevier, Cambridge, MA, EUA, 2nd edition.
- Schrepp, M., Hinderks, A., and Thomaschewski, J. (2017). Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(4):40–44.
- Schrepp, M. and Thomaschewski, J. (2019). Design and Validation of a Framework for the Creation of User Experience Questionnaires. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(7):88–95.
- Silva, G. R., Donat, J. C., Rigoli, M. M., de Oliveira, F. R., and Kristensen, C. H. (2016). A questionnaire for measuring presence in virtual environments: factor analysis of the presence questionnaire and adaptation into Brazilian Portuguese. *Virtual Reality*, 20(4):237–242.
- Tcha-Tokey, K., Christmann, O., Loup-Escande, E., and Richir, S. (2016). Proposition and Validation of a Questionnaire to Measure the User Experience in Immersive Virtual Environments. *International Journal of Virtual Reality*, 16(1):33–48.
- Tcha-Tokey, K., Loup-Escande, E., Christmann, O., Canac, G., Farin, F., and Richir, S. (2018). Towards a user experience model in immersive virtual environments. *Advances in Human-Computer Interaction*, 2018:1–10.

Vermeeren, A., Law, E., Roto, V., Obrist, M., Hoonhout, J., and Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: Current state and development needs. In *Proceedings of the Nordic Conference on Human-Computer Interaction*, pages 521–530, New York, NY, EUA. ACM.