



Desenvolvimento do método para avaliação de aplicativos móveis em saúde

Development of a method for evaluating mobile health applications

Desarrollo de un método para evaluar aplicaciones móviles de salud

Thais Maria Yomoto Ferauche¹, Jean José Silva², Marcia Ito¹

¹ Ma. e Dra. Programa de Mestrado Profissional em Sistemas Produtivos do Centro Paula Souza, UEPEP/CEETEPS, São Paulo (SP), Brasil.

² Dr., Departamento, Santa Casa de Misericórdia de Passos, Passos (MG), Brasil.

{thais.ferauche, marcia.ito}@cpspos.sp.gov.br, jeanjsilva.usp@gmail.com

Resumo

Objetivo: A crescente disponibilidade de aplicações de saúde destaca a necessidade de garantir informações confiáveis, promovam a adesão ao tratamento recomendado pelos profissionais de saúde e previnam riscos decorrentes de informações imprecisas ou inadequadas. No entanto, a falta de métodos para avaliar essas aplicações dificulta a seleção. Um método foi desenvolvido baseado em literatura, necessidades de usuários e profissionais, melhores práticas e evidências científicas. **Método:** O método combina revisão sistemática da literatura e desenvolvimento experimental, manifestado na forma de questionário, validado por especialistas. **Resultados:** Desenvolvimento do método de avaliação para aplicativos de saúde. **Conclusão:** Pequenos ajustes foram feitos. A análise Kappa de Fleiss indica sua reprodutibilidade satisfatória.

Descritores: Avaliação da Tecnologia Biomédica; Saúde Digital; Segurança do Paciente

Abstract

Objective: The increasing availability of health applications highlights the need to ensure reliable information, promote adherence to treatment recommended by health professionals and prevent risks arising from inaccurate or inadequate information. However, the lack of methods to evaluate these applications makes selection difficult. A method was developed based on literature, user and professional needs, best practices and scientific evidence. **Method:** The method combines systematic



literature review and experimental development, manifested in the form of a questionnaire, validated by experts. Results: Development of the evaluation method for health applications. Conclusion: Small adjustments were made. Fleiss' Kappa analysis indicates satisfactory reproducibility.

Keywords: *Assessment of Biomedical Technology; Digital Health; Patient safety*

Resumen

Objetivo: La creciente disponibilidad de aplicaciones de salud pone de relieve la necesidad de garantizar información confiable, promover la adherencia al tratamiento recomendado por los profesionales de la salud y prevenir riesgos derivados de información inexacta o inadecuada. Sin embargo, la falta de métodos para evaluar estas aplicaciones dificulta la selección. Se desarrolló un método basado en la literatura, las necesidades de los usuarios y profesionales, las mejores prácticas y la evidencia científica. Método: El método combina la revisión sistemática de la literatura y el desarrollo experimental, manifestado en forma de cuestionario, validado por expertos. Resultados: Desarrollo del método de evaluación de aplicaciones en salud. Conclusión: Se hicieron pequeños ajustes. El análisis Kappa de Fleiss indica una reproducibilidad satisfactoria.

Descriptores: *Evaluación de Tecnología Biomédica; Salud Digital; Seguridad del paciente.*

Introdução

Os aplicativos móveis de saúde oferecem acesso a informações, monitoramento de condições e apoio ao autocuidado. A ausência de critérios claros de avaliação dificulta a seleção de aplicativos confiáveis, afetando cidadãos, pacientes, cuidadores e profissionais de saúde. A falta de métodos de avaliação específicos e a análise insuficiente de aspectos como acessibilidade e segurança de dados representam desafios adicionais^{(1) (2)}.

A Avaliação de Tecnologia em Saúde (*Health Technology Assessment – HTA*) é um processo sistemático que avalia a eficácia, segurança e custo das tecnologias de saúde, auxiliando em decisões políticas e melhorando desfechos clínicos. Esse



processo aplica-se a diversos contextos, de medicamentos a dispositivos médicos, e diversos métodos foram desenvolvidos para garantir a qualidade dos aplicativos de saúde^{(3) (4)}.

Dos métodos existentes oito deles são os mais utilizados e conhecidos e que são específicos para a área da saúde: Escala de Classificação de Aplicativos para Dispositivos Móveis (MARS), Escala de Classificação de Aplicativos para Dispositivos Móveis Adaptado (A-MARS), Escala de Classificação de Aplicativos para Dispositivos Móveis – Versão do usuário (uMARS), Escala de Usabilidade do Sistema (SUS), Questionário de Usabilidade de Aplicativos para saúde (MAUQ), Questionário de Usabilidade de Aplicativos de Saúde (PSSUQ), Relatório e Avaliação de Evidências de Saúde Móvel (mERA - OMS) e ISO/TS 82304-2⁽⁵⁾.

Apesar da variedade de métodos, nenhum cobre todos os aspectos necessários. Um método abrangente pode proporcionar uma avaliação mais completa dessas ferramentas ⁽⁵⁾.

Desta forma, foi desenvolvido um método de avaliação de aplicativos de saúde para garantir a segurança do paciente. O método avalia a confiabilidade das informações, confiabilidade das informações, além de outros critérios, como acessibilidade, adequação, usabilidade, segurança de dados, utilidade, interoperabilidade e infraestrutura. Oferece diretrizes para selecionar aplicativos seguros e confiáveis na área da saúde, integráveis a outros sistemas de informação a fim de melhorar o cuidado dos pacientes.

A principal contribuição deste estudo se encontra no desenvolvimento e validação de um método de avaliação para aplicativos móveis de saúde, com foco na confiabilidade das informações de saúde. Esse método, após as validações, estará disponível publicamente e permitirá indicar os aplicativos que possuem as características mínimas de confiabilidade. Como resultado, espera-se que o método desenvolvido contribua para uma avaliação mais segura e confiável de aplicativos de saúde, desempenhando um papel na promoção da saúde, em consonância com os desafios dos sistemas de informação aplicados à área da saúde.

Métodos

A pesquisa adotou uma abordagem quali-quantitativa com uma posição epistemológica interpretativista. Destacou-se a ênfase na compreensão das



perspectivas e experiências dos participantes em relação ao protocolo de avaliação de aplicativos na área de saúde. O desenvolvimento do método de avaliação para aplicativos de saúde (protocolo) derivou da identificação dos métodos de avaliação para aplicativos na área da saúde existentes e suas características encontrados na revisão sistemática da literatura de Ferauche, Silva e Ito⁽⁵⁾, que foi baseando-se na ISO/IEC 25000⁽⁶⁾.

O desenvolvimento do protocolo, pautado na ISO/IEC 25000, foi dividido em etapas, incluindo o segundo processo da norma ISO/IEC 25000 (avaliação da qualidade). As etapas compreenderam: (a) definição dos objetos e contextualização da avaliação do protocolo que define os objetivos da avaliação, identificação das áreas, definição das categorias; (b) definição e configuração do modelo de qualidade do protocolo que é um conjunto de critérios e requisitos com embasamento na literatura e critérios restritivos; (c) identificação das medidas da avaliação do protocolo que são os requisitos mapeados em atributos que podem ser medidos; (d) definição de níveis de pontuação e critérios de julgamento do protocolo que é a especificação dos tipos de respostas e atribuição da pontuação; (e) definição da avaliação do protocolo que é a elaboração da avaliação a ser aplicada; (f) execução da avaliação do protocolo que compreendem: (i) coletar as medidas; (ii) comparar as medidas; e (iii) avaliar os dados obtidos⁽⁶⁾.

Desta forma, o plano de validação do protocolo inclui as diretrizes, procedimentos necessários e o método para coletar os dados. No caso deste protocolo, a técnica utilizada para a validação é a entrevista. A validação por entrevista é uma técnica usada em pesquisas qualitativas para obter credibilidade. Na etapa conclusiva da validação, o avaliador tem três responsabilidades chave: (i) reunir as medidas de qualidade; (ii) contrastar as medidas com os critérios de qualidade pré-estabelecidos; e (iii) analisar os dados recolhidos durante a avaliação⁽⁷⁾.

As análises da validação do protocolo foram as qualitativas para ajustes no protocolo e quantitativas para avaliar a reprodutibilidade do instrumento. As técnicas de análise de dados adotadas são a análise qualitativa do conteúdo, auxiliando na identificação de padrões e tendências nas respostas dos participantes e na



compreensão do tema em questão e, na análise quantitativa de Kappa de Fleiss, para avaliar a reprodutibilidade do instrumento proposto.

Resultados e Discussão

Um mapeamento sistemático, conforme Ferauche, Silva e Ito⁽⁵⁾, identificou características relevantes em aplicativos de saúde, formando a estrutura inicial para o método de avaliação. Pesquisas adicionais, como as conduzidas por Stoyanov *et al.*^{(8) (9)} e Roberts *et al.*⁽¹⁰⁾ orientaram os estágios do processo de desenvolvimento do método de avaliação. O protocolo foi baseado na ISO/IEC 25000⁽⁶⁾, unificando as normas ISO/IEC 9126⁽¹¹⁾ e ISO/IEC 14598⁽¹²⁾ para garantir a qualidade do software, envolvendo especificação de requisitos e avaliação da qualidade.

Nesta pesquisa, para a especificação dos requisitos, o propósito estabelecido para o método é a análise e comparação de aplicativos de saúde, visando avaliar a sua qualidade e confiabilidade por profissionais do setor de saúde. Assim os requisitos de qualidade de um software são convertidos em características mensuráveis para identificar as medidas de avaliação. Em resumo, o método de avaliação, denominado QRAMHe, é elaborado por meio do desenvolvimento de questões simples e objetivas, organizadas em um questionário, com respostas nominais definidas para avaliar o grau de qualidade e confiabilidade do aplicativo. Esse método está disponível publicamente por meio de uma ferramenta computacional (<https://qramhe.org/>).

O principal componente do QRAMHe “Definição dos Objetos e Contextualização” é o seu modelo de qualidade e confiabilidade, responsável por definir o conjunto de características que fornece a base para a avaliação de qualidade e confiabilidade de aplicativos na área da saúde.

O modelo do método QRAMHe é dividido em quatro grandes áreas: fornecedor, adoção da tecnologia, saúde e qualidade subjetiva. Além das grandes áreas, o método possui uma estrutura hierárquica composta por categorias. Essa divisão foi fundamentada pela análise da revisão da literatura.

Em Fornecedor, preocupou-se em avaliar as informações sem que houvesse a necessidade de contactar o fornecedor do aplicativo, como acontece em alguns métodos estudados. Portanto, foram consideradas as informações iniciais sobre o aplicativo, atualização e descrição do aplicativo e plataformas de distribuição.



Na Adoção da Tecnologia, enfatiza-se a facilidade de uso, interface amigável e funcionalidades claras do aplicativo. Os usuários devem navegar e utilizar o aplicativo eficientemente, sem dificuldades desnecessárias. É essencial verificar se o fornecedor assegura a segurança e privacidade dos dados dos usuários, incluindo medidas como criptografia e conformidade com regulamentos de privacidade, apresentadas de forma transparente.

Em Saúde, são definidos quatro domínios: adequação e restrição, integração, utilidade e funcionalidade, e confiabilidade nas informações. A integração visa facilitar a troca de informações entre o aplicativo e outros sistemas ou profissionais de saúde, garantindo segurança ao chamar funcionalidades de outros sistemas. Em utilidade e funcionalidade é utilizado para verificar se o aplicativo atende a meta estabelecida e se há restrições, incluindo compensações monetárias. Para garantir a autenticidade das informações no aplicativo, o último domínio tem como meta avaliar a confiabilidade das informações.

Por fim, a Qualidade Subjetiva, refere-se à percepção individual de qualidade, variando de pessoa para pessoa com base em opiniões, preferências e experiências pessoais. Indivíduos podem interpretar e avaliar a qualidade de algo de forma diferente, de acordo com suas expectativas, valores e critérios pessoais. Esta medida subjetiva é aplicável em várias áreas, incluindo produtos, serviços e experiências. É importante reconhecer que a qualidade subjetiva varia conforme a perspectiva e contexto de cada indivíduo.

Em sequência é apresentado cada uma das categorias originadas pelas quatro áreas.

A categoria 'Informações Iniciais' avalia a disponibilidade das informações essenciais nos aplicativos, como detalhes de contato (site e e-mail) para suporte e mais informações. Verifica-se a presença de uma política de privacidade para proteger os dados dos usuários e a facilidade de localização e acesso ao aplicativo. Esses aspectos garantem transparência, confiabilidade e acessibilidade aos potenciais usuários.

Na categoria 'Adequação e Restrição', avalia-se se o aplicativo satisfaz às necessidades do usuário, considerando restrições de idade ou classificação



indicativa, diagnóstico inicial e possíveis limitações do aplicativo, como restrições de uso ou falta de funcionalidades.

A categoria 'Usabilidade' verifica se o aplicativo apresenta uma interface intuitiva, amigável e de fácil compreensão, permitindo o uso eficiente por profissionais de saúde. Isso inclui a organização lógica dos elementos, uso de ícones e símbolos claros, e disposição das informações de forma compreensível.

A 'Segurança e Privacidade' dos dados é analisada para garantir que o aplicativo adote medidas de segurança, como criptografia e armazenamento seguro, e esteja em conformidade com regulamentos de privacidade, como a LGPD, GDPR e HIPAA, obtendo consentimento dos usuários e garantindo o uso adequado das informações.

Na categoria 'Utilidade e Funcionalidade', verifica-se se o aplicativo cumpre sua proposta, fornecendo funcionalidades que atendem às necessidades dos usuários e avaliando possíveis restrições de uso associadas a pagamentos, como assinaturas ou compras no aplicativo. Avalia-se também o custo-benefício e a transparência das restrições.

A categoria 'Integração com Sistemas de Saúde e Profissionais' avalia a capacidade do aplicativo de se conectar aos sistemas dos profissionais de saúde, facilitando a troca de informações e garantindo uma boa comunicação e compreensão do quadro clínico do paciente.

A 'Confiabilidade das Informações' verifica se o aplicativo fornece informações respaldadas por evidências científicas e fontes confiáveis, atualizadas de acordo com as diretrizes médicas, e se cita suas referências de forma apropriada.

A categoria 'Infraestrutura' abrange aspectos essenciais para o funcionamento adequado e seguro do aplicativo, como a segurança das informações dos pacientes, capacidade de sincronização e backup dos dados, integração com sistemas de comunicação de terceiros, proteção de dados pessoais e compatibilidade com diferentes dispositivos.

A categoria 'Acessibilidade' garante que o aplicativo seja utilizável por todas as pessoas, independentemente de suas capacidades físicas ou sensoriais, promovendo a inclusão e igualdade de acesso aos serviços de saúde digital.



A categoria 'Qualidade Subjetiva' refere-se à satisfação pessoal do usuário com o aplicativo, considerando fatores como facilidade de uso, design da interface, clareza das informações e capacidade de atender às necessidades do usuário. A satisfação do usuário reflete o quanto o aplicativo atende às suas expectativas e pode incentivar o uso contínuo.

O componente "Definição e Configuração do Modelo de Qualidade" permite uma compreensão mais precisa do desempenho do aplicativo de saúde. As medidas, conforme as diretrizes da ISO/IEC 25000⁽⁶⁾, mapeiam as categorias em atributos mensuráveis, usando questionários para coletar dados. Assim, foi desenvolvido o questionário QRAMHe que usou como base os diversos métodos de avaliação mencionados. As questões contidas nesse questionário foram adaptadas para o contexto de aplicativos da área da saúde. Além disso, o modelo de qualidade e confiabilidade descrito anteriormente foi usado para a criação de novas perguntas. O questionário final é composto por 50 questões, abordando as 10 categorias definidas no modelo. Para as respostas, seguindo as diretrizes da área da saúde neste tema, optou-se por utilizar a Escala de Likert nas questões de avaliação do aplicativo⁽¹³⁾. No protocolo, a escala varia com 5 alternativas/pontos. Os respondentes são solicitados a escolher o nível de concordância ou discordância que melhor representam sua opinião em relação a cada afirmação apresentada.

Para o componente "Identificação das Medidas da Avaliação do Protocolo", ao estabelecer critérios de avaliação, é possível determinar o nível de qualidade e confiabilidade de um aplicativo. Para definir tal nível de qualidade e confiabilidade do aplicativo, é necessário atribuir valores quantitativos para as perguntas contidas no questionário. Por esse motivo, as respostas obtidas na aplicação da avaliação de qualidade e confiabilidade devem ser convertidas em dados quantitativos. As pontuações são atribuídas a cada pergunta do questionário. As respostas do questionário são de escolha única, variando de 1 a 5, na qual 1 representa um nível muito baixo naquele quesito e, no outro extremo, a resposta 5 representa um nível muito elevado. Cada questão foi elaborada, incluindo diferentes opções de resposta, a fim de fornecer uma compreensão das avaliações.

O componente "Definição de Níveis de Pontuação e Critérios de Julgamento do Protocolo" estabelecer padrões e objetivos para avaliar e pontuar os diferentes



aspectos no protocolo de avaliação. O cálculo final é definido pela média aritmética da pontuação de cada categoria. Obtendo um valor de 1 a 5, é possível identificar a classificação para cada categoria: sendo que para valores até 1 ponto, fica classificado que a categoria é ineficiente; valores de 1 a 2, a categoria é classificada deficiente; valores de 2 a 3, a categoria é classificada como regular; para valores de 3 a 4, a categoria é classificada como aceitável; e, para valores de 4 a 5, a categoria é classificada como eficiente. A partir dos resultados obtidos para cada categoria, o cálculo para obter o nível de qualidade geral do aplicativo é dado pela média aritmética das categorias. Para definir o nível de qualidade, o resultado é dividido em três com seus respectivos valores padrões: (i) Superior, pontuação igual ou maior que 80%; (ii) Médio, pontuação igual ou maior que 50% e menor que 80%; e (iii) Baixo, pontuação menor que 50%.

De acordo com os possíveis níveis de qualidade do aplicativo, considera-se que um aplicativo com nível superior possui um ótimo nível de qualidade. Assim, o profissional pode usar e indicar o aplicativo sem receio. Por outro lado, um nível de qualidade médio revela que o aplicativo pode ser indicado para uso com restrições. Por fim, um nível de qualidade baixo sugere que o aplicativo não conseguiu alcançar níveis satisfatórios de qualidade. Dessa maneira, o resultado da avaliação sugere que esse aplicativo não deva ser adotado ou indicado para uso pelo usuário do aplicativo.

Para a avaliação inicial do protocolo do método, optou-se pela entrevista com especialistas da área, antes de avaliá-la com uma experimentação de campo. A validação por entrevista é uma técnica usada em pesquisas qualitativas para obter credibilidade(13). Esta pesquisa foi aprovada pelo Comissão de Ética do Programa de Mestrado Profissional em Gestão e Tecnologia em Sistemas Produtivos.

O planejamento foi adaptado de Mann⁽¹⁴⁾ e foram necessárias as etapas: (i) objetivo; (ii) participantes; (iii) preparação; (iv) realização; e (v) pós-entrevista. O objetivo da entrevista é validar o protocolo QRAMHe. Os três participantes selecionados são doutores com diversas especialidades que combinam conhecimentos em aplicativos móveis, tecnologia e saúde. Para a condução do experimento foram agendadas entrevistas com duração entre 30 e 60 minutos e conduzidas à distância por meio de videoconferência. Para cada entrevistado, a



entrevista semiestruturada foi guiada por meio de um questionário composto de dez perguntas.

Os entrevistados foram indagados sobre a clareza e compreensão do questionário do protocolo. As opiniões se dividiram: enquanto alguns acharam as perguntas fáceis de entender, outros apontaram que as respostas eram longas e repetitivas, sugerindo que poderiam ser mais concisas. Além disso, a estrutura das perguntas, especialmente aquelas com múltiplas subperguntas, pode complicar as respostas.

Os participantes discutiram a importância da avaliação de aplicativos de saúde, concordando que as perguntas são pertinentes tanto para a área de saúde quanto para a tecnologia. Destacou-se a necessidade de usabilidade e acessibilidade, embora tenha havido preocupação com a abrangência do questionário. A relevância das perguntas foi confirmada, com sugestões para melhorar a terminologia e incluir questões sobre integração de sistemas de informações em saúde.

Os entrevistados opinaram sobre a adequação das respostas no questionário para avaliar os aplicativos de saúde, com sugestões para melhorias. Consideraram as respostas apropriadas, mas recomendaram otimizar a organização das opções. Propôs-se colocar a resposta principal no início para maior eficiência. A variação nas respostas foi notada, sugerindo simplificar para tornar o questionário mais amigável. Um entrevistado achou a escala de cinco pontos adequada, permitindo a granularidade necessária na avaliação.

Os entrevistados foram questionados sobre a abrangência do questionário em áreas como acessibilidade, adequação, usabilidade, segurança, privacidade, utilidade, integração com sistemas de saúde e confiabilidade das informações, consideradas relevantes para avaliar aplicativos de saúde. Eles concordaram com a relevância dessas áreas. No entanto, expressaram preocupações sobre a especificidade em relação à saúde, sugerindo maior clareza para os profissionais de saúde que avaliam os aplicativos para seus pacientes.

As opiniões dos entrevistados sobre as perguntas do questionário são favoráveis, pois abordam os principais aspectos de avaliação de aplicativos na área de saúde. No entanto, há preocupações com a clareza da associação entre as



categorias e o objetivo principal do instrumento de avaliação. Além disso, um entrevistado sugere incluir uma pergunta sobre a participação dos aplicativos em estudos clínicos randomizados, reconhecendo sua relevância para profissionais de saúde, embora possa ser desafiador obter essa informação.

Quanto à organização do questionário, as opiniões dos entrevistados variam. Alguns elogiam a sequência lógica e a divisão por etapas, enquanto outros expressam preocupações sobre a clareza das categorias. Uma entrevistada sugere que profissionais de saúde podem priorizar a utilidade e funcionalidade do aplicativo. A segurança é reconhecida como importante, mas há a necessidade de enfatizar essa área, especialmente em relação à privacidade de dados.

Os entrevistados consideram o questionário abrangente para avaliar aplicativos na área de saúde, com a meta na avaliação dos aplicativos em si. No entanto, reconhecem a complexidade da área da saúde e a possibilidade de questões genéricas não abordarem todas as especificidades. Sugerem que profissionais de saúde possam desenvolver questões específicas para determinadas áreas de saúde, se necessário. Uma entrevistada sugere incluir uma pergunta sobre integração com dispositivos vestíveis, considerando sua relevância para muitos aplicativos de saúde.

Os entrevistados tiveram opiniões diversas quando questionados sobre possíveis dificuldades em responder ao questionário. Um deles mencionou que termos técnicos como "usabilidade" podem ser desafiadores para pessoas sem conhecimento técnico, mas reconheceu sua importância na avaliação dos aplicativos. Portanto, não sugeriu mudanças na terminologia. Os outros dois entrevistados, familiarizados com a área, não encontraram dificuldades e consideraram o questionário adequado em termos de clareza e terminologia.

Quando questionados sobre o tempo estimado para concluir o questionário, os entrevistados apresentaram perspectivas divergentes. Alguns consideram que o tempo previsto de 20 minutos é insuficiente para fornecer respostas precisas e sugeriram que seriam necessários pelo menos 30 minutos para uma avaliação minuciosa. Outros acreditam que 30 minutos é razoável, mas também reconhecem que responder em 30 minutos pode ser suficiente, dependendo se o avaliador está



agindo como um profissional de saúde ou um usuário final do aplicativo, o que pode influenciar a rapidez do processo.

Com base nas respostas dos entrevistados, diversas sugestões específicas para a melhoria do questionário foram apresentadas. Essas incluem substituir a escala Likert por *checkboxes* em algumas questões, dividir questões, ajustar terminologia, adicionar perguntas específicas, fornecer informações sobre o objetivo de cada categoria, tornar as perguntas mais sintéticas, focar a confiabilidade dos dados e considerar a situação de aplicativos novos.

Para avaliar a reprodutibilidade do instrumento proposto, protocolo QRAMHe, foi elaborado um questionário composto por dez itens, indagando sobre os fatores: (1) entendimento das questões, (2) relevância, (3) perguntas adequadas, (4) abrangência, (5) aspectos principais, (6) organização, (7) aplicabilidade do aplicativo, (8) dificuldade em responder, (9) tempo de resposta e (10) melhoramentos.

Foi utilizada a estatística de Kappa de Fleiss com o auxílio do software IBM SPSS Statistics para Windows (versão 28.0) para avaliar a confiabilidade das respostas dos entrevistados. O Kappa de Fleiss é uma medida estatística utilizada para avaliar a confiabilidade e concordância entre vários avaliadores⁽¹⁵⁾. Portanto, foi possível avaliar a reprodutibilidade do questionário do protocolo QRAMHe. O nível de significância adotado foi de $\alpha = 0,05$. O teste Kappa de Fleiss identificou uma confiabilidade geral razoável significativa entre os três avaliadores ($k = 0,298$ [IC 95%: 0,011-0,586]; $z = 2,032$; $p < 0,043$). No entanto, para a resposta "SIM", o teste identificou confiabilidade moderada ($k = 0,441$ [IC 95%: 0,083-0,799]; $z = 2,415$; $p < 0,017$).

A análise Kappa de Fleiss revelou uma reprodutibilidade significativa ($p < 0,017$) de acordo com as respostas dos três avaliadores. Isso indica que o instrumento é satisfatoriamente reprodutível.

Conclusão

Esta pesquisa aborda um problema relacionado à falta de métodos de avaliação abrangentes que englobem critérios específicos para aplicativos na área de saúde. Os objetivos desta pesquisa incluíram a criação de um método de



avaliação específico para aplicativos móveis na área de saúde, visando avaliar critérios como acessibilidade, adequação, usabilidade, segurança e privacidade, utilidade, funcionalidade, interoperabilidade, confiabilidade das informações de saúde e infraestrutura.

A principal contribuição deste estudo reside no desenvolvimento de um método de avaliação aberto para aplicativos móveis de saúde, com foco na confiabilidade das informações de saúde. Esse método está disponível publicamente por meio de uma ferramenta computacional (<https://qramhe.org/>) e poderá avaliar aplicativos na área da saúde, auxiliando na seleção por profissionais de saúde de aplicativos confiáveis para uso de pacientes e cuidadores.

A análise das respostas dos entrevistados referentes ao protocolo inicial proporciona uma visão sobre a eficácia desse instrumento. As diversas opiniões dos participantes destacam a necessidade de aprimorar a clareza, concisão e relevância do questionário. Fica evidente a complexidade na avaliação de aplicativos de saúde, considerando as variadas necessidades de públicos diversos, desde profissionais de saúde até leigos na área. As sugestões dos entrevistados apontam na direção promissora do desenvolvimento de um questionário mais acessível e eficaz, com o potencial de auxiliar na avaliação dos aplicativos na área da saúde. Além disso, na análise que o protocolo é satisfatoriamente reproduzível.

Este modelo de avaliação almeja facilitar a seleção de aplicativos de saúde confiáveis, incentivando uma seleção criteriosa tanto por parte dos usuários como dos profissionais de saúde. Sua aplicação visa contribuir para verificar a segurança, efetividade e confiabilidade das informações médicas fornecidas por aplicativos móveis, resultando na indicação por parte dos profissionais de saúde de aplicativos que possam trazer experiências satisfatórias e desfechos clínicos desejados.

Esses resultados destacam a necessidade de ajustes no questionário para a conclusão da fase de validação do protocolo. Pretende-se, como trabalho futuro, testar a avaliação com um experimento de campo na qual os profissionais de saúde irão avaliar aplicativos reais e, assim, verificar sua eficácia em relação às características definidas.

Em suma, este estudo oferece um método de avaliação para aplicativos móveis na área de saúde, abordando as complexidades inerentes à confiabilidade



das informações de saúde. Um instrumento como esse pode impedir que se proliferem ou que os usuários utilizem aplicativos poucos confiáveis e que podem inclusive prejudicar a saúde de quem o usa.

Referências

1. Choi W, Wang S, Lee Y, Oh H, Zheng Z. A systematic review of mobile health technologies to support self-management of concurrent diabetes and hypertension. *Journal of the American Medical Informatics Association*. 1º de junho de 2020;27(6):939–45.
2. Agarwal P, Gordon D, Griffith J, Kithulegoda N, Witteman HO, Sacha Bhatia R, *et al.* Assessing the quality of mobile applications in chronic disease management: a scoping review. *npj Digit Med*. 10 de março de 2021;4(1):46.
3. Wiethoff I, Evers SMAA, Michels M, Hiligsmann M. An introduction to health technology assessment and health economic evaluation: an online self-learning course. *Neth Heart J*. 1º de junho de 2023;31(6):219–25.
4. Garfield S, Armstrong S, Nguyen J, Hildreth C, Wu B, James I, *et al.* Are Health Technology Assessments Keeping Pace with Health Equity Priorities: A Review of Existing Approaches and Discussion of Emerging Practices. *Medical Research Archives [Internet]*. 26 de junho de 2023 [citado 5 de novembro de 2023];11(6). Disponível em: <https://esmed.org/MRA/mra/article/view/3903>
5. Ferauche TMY, Silva JJ, Ito M. Proposal for the Protocol for Assessment of mHealth: QRAMHe. *CONTECSI - International Conference on Information Systems and Technology Management [Internet]*. 2024;(20). Disponível em: <https://www.tecsi.org/contecsi/index.php/contecsi/20thCONTECSI/paper/view/7213>
6. ISO 25000. ISO. 2014 [citado 27 de outubro de 2023]. ISO/IEC 25000:2014. Disponível em: <https://www.iso.org/standard/64764.html>
7. Simpson SA, McDowell AK. Validate and Move: Use Validation as a Transitional Tool in the unwieldy interview [Internet]. [citado 15 de outubro de 2023]. Disponível em: <https://www.taylorfrancis.com/chapters/edit/10.4324/9780429437243-23/validate-move-scott-simpson-anna-mcdowell>
8. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps. *JMIR mHealth uHealth*. 11 de março de 2015;3(1):e27.
9. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). *JMIR Mhealth Uhealth*. 10 de junho de 2016;4(2):e72.



10. Roberts AE, Davenport TA, Wong T, Moon HW, Hickie IB, LaMonica HM. Evaluating the quality and safety of health-related apps and e-tools: Adapting the Mobile App Rating Scale and developing a quality assurance protocol. *Internet Interventions*. abril de 2021;24:100379.
11. ISO 9126. ISO. 2001 [citado 15 de outubro de 2023]. ISO/IEC 9126-1:2001. Disponível em: <https://www.iso.org/standard/22749.html>
12. ISO 14598. ISO. 1998 [citado 15 de outubro de 2023]. ISO/IEC 14598-5:1998. Disponível em: <https://www.iso.org/standard/24906.html>
13. Barua A. Methods for Decision-Making in Survey Questionnaires Based on Likert Scale. *Journal of Asian Scientific Research*. 1º de janeiro de 2013;3(1):35–8.
14. Mann S. *The Research Interview* [Internet]. London: Palgrave Macmillan UK; 2016 [citado 15 de outubro de 2023]. Disponível em: <http://link.springer.com/10.1057/9781137353368>
15. Gwet KL. Large-Sample Variance of Fleiss Generalized Kappa. *Educational and Psychological Measurement*. 1º de agosto de 2021;81(4):781–90.