

Capítulo

3

Avaliação de Tecnologia em Saúde Digital: Fundamentos e Aplicações Práticas

Thais Maria Yomoto Ferauche (UEPEP/CEETEPS), Jean José Silva (Santa Casa de Misericórdia de Passos), Wilian Hisatugu (UFES), Márcia Ito (UEPEP/CEETEPS)

Abstract

The use of digital technologies in healthcare has grown significantly, both in support of public and private healthcare. These technologies have shown considerable potential for improving healthcare services. In this scenario, it is necessary that the assessment of health technologies incorporate specific processes to assess the use of these digital technologies. The purpose of this course is to present the relevance, fundamentals and applications of technology assessment in digital health, with a particular focus on the assessment of software for the healthcare. The particularities of applications in the healthcare will be addressed, as well as specific protocols and processes for evaluating digital health technologies.

Resumo

O emprego crescente de tecnologias digitais na área da saúde tem sido expressivo, tanto no apoio aos serviços de saúde pública quanto privada. Estas ferramentas têm evidenciado um considerável potencial para aprimorar o suporte oferecido pelos serviços de saúde. Nesse cenário, é necessário que a avaliação de tecnologias em saúde incorpore processos específicos para avaliar as tecnologias digitais. O propósito deste minicurso é apresentar a relevância, os fundamentos e as aplicações da avaliação de tecnologia em saúde digital, com um foco particular na avaliação de softwares para a área da saúde. Serão abordadas as particularidades dos aplicativos na área da saúde, bem como os protocolos e processos específicos para a avaliação de tecnologias digitais na saúde.

3.1 Introdução

A evolução da tecnologia da informação e comunicação (TIC) aliada a inteligência artificial, ao IoT (Internet das coisas) e a robótica fizeram com que os processos se tornassem integrados e inteligentes desde a produção da matéria prima até o produto, permitindo aumentar a produtividade, a criação de valor e o bem-estar social. Na Indústria, a transformação digital, alterou a estrutura organizacional, fazendo com ela seja mais ágil e responsiva. Assim surge a Indústria 4.0, na qual criam-se papéis que encorajam a inovação e a mudança de cultura aumentando a colaboração e a experimentação entre desenvolvedores e clientes. A mudança é tão significativa que altera o modo de interação dos clientes com a organização de serviços por meio de canais digitais. O impacto da transformação digital nas indústrias inclui o aumento da produtividade na manufatura, aumento da qualidade e eficiência na assistência e indústria de saúde, melhora o acesso à educação na medida que novos meios de disseminação do conhecimento são desenvolvidos, melhoram as condições de trabalhos, facilitam a tomada de decisão pelas organizações e por fim estimulam o crescimento econômico por meio das cidades inteligentes [ABIODUN; RAMPERSAD; BRINKWORTH, 2023].

Portanto, estamos numa nova era denominada, Era Digital, na qual tecnologia, máquinas e seres humanos estão conectados e interagindo em tempo real e em muitas vezes sendo difícil perceber quem é quem na execução do serviço. Na Era digital a globalização e a rápida evolução das tecnologias digitais estão trazendo mudanças diversas e complexas para a sociedade. Ao mesmo tempo, o mundo está se deparando com desafios em escala global, como o esgotamento dos recursos naturais, o aquecimento global, o terrorismo entre outros. Em setembro de 2015 a ONU estabeleceu a agenda 2030 para o desenvolvimento sustentável, o desafio é estabelecer um sistema em que todas as nações trabalhem juntas com a finalidade de ter um mundo mais sustentável, resolvendo tanto o desenvolvimento econômico quanto as questões sociais. A chave para isto é que pessoas e o planeta tenham paz e prosperidade, com uma inclusão para todos sem exceção. Foi assim que em janeiro de 2016, o Japão define a Sociedade 5.0, a sociedade centrada no humano, na qual tanto o desenvolvimento social quanto a solução dos desafios sociais são alcançados, e todos possam desfrutar de uma boa qualidade de vida sem diferenças socioeconômicas. Para isto é preciso equilibrar o avanço econômico com a resolução de problemas sociais na qual computadores e hiper conexão resultem em um modo de vida mais inteligente, eficiente e sustentável [SHEPHERD, 2004] [FUKUYAMA, 2018].

As transformações que a indústria 4.0 e as metas que se deseja alcançar na sociedade 5.0, afetam todas as áreas inclusive a da saúde e com isso faz surgir o conceito de Saúde Digital que expande o conceito da e-Saúde ao incluir os consumidores digitais que possuem os mais diversos dispositivos conectados e inteligentes (WHO, 2021).

Em anos mais recentes, notou-se que para melhorar a qualidade de vida e por consequência a saúde do cidadão, era preciso cuidar não somente da doença, mas do paciente como um todo, seu estado psicossocial, econômico, entre outros. A este tipo de assistência é denominado de cuidado centrado no paciente e a indústria de saúde passa a ter um “consumidor de serviço de saúde” que deseja e tem o controle sobre a gestão da sua saúde. É assim que a indústria de saúde passa a se preocupar com as necessidades e expectativas com este novo “consumidor de serviço de saúde” que busca por experiências personalizadas, confortáveis, rapidez e eficiência na prestação do serviço. A saúde digital

além de melhorar os serviços de saúde trazendo agilidade e o aumento de produtividade, possibilita fornecer ao consumidor de serviço de saúde a experiência inovadora e o controle e gestão sobre sua saúde [ITO, 2020] [STOUMPOS; KITSIOS; TALIAS, 2023]. Assim, surge o consumidor de saúde digital que são os cidadãos, pacientes, cuidadores e familiares que usam os serviços e ferramentas da saúde digital. Os serviços e ferramentas da saúde digital são os que permitem o consumidor de saúde digital ter um papel ativo em sua saúde e na prevenção de doenças de forma segura. A diferença com relação à telessaúde é que neste caso o usuário obrigatoriamente deve ter um papel ativo na condução de sua saúde. Outro ponto é que ele não pode ser uma solução para uso somente do profissional de saúde [WETTER, 2012].

Por outro lado, com a erradicação das doenças infecciosas, houve um aumento das doenças crônicas. Há diferença de abordagem entre os dois tipos de doença, enquanto na doença infecciosa a dependência da participação do paciente em relação ao seu tratamento é menor (não é nulo, pois decidir se vai tomar ou não o medicamento é do paciente), na doença crônica é totalmente dependente do doente, pois a decisão de parar de fumar ou mudar os hábitos alimentares é dele e essa decisão implicará num bom desfecho clínico ou não. Assim, um dos fatores de sucesso no tratamento de doenças é a participação ativa do paciente em relação ao seu estado de saúde. Neste sentido, a saúde digital contribui ao facilitar a participação dos pacientes no processo de tomada de decisão quanto ao seu estado de saúde. O paciente ao buscar por informações seja pela internet ou por aplicativos de saúde digital, por exemplo, aplicativos em celulares, fazem a diferença na tomada de decisão por parte deles. Pode-se então dizer que os consumidores de saúde digital se beneficiam com serviços e ferramentas digitais que possuem informações e funcionalidades que os ajudam na tomada de decisão sobre sua saúde, trazendo um empoderamento para estes cidadãos [ITO, 2020] [ITO; PEREIRA, 2015] [STOUMPOS; KITSIOS; TALIAS, 2023].

De acordo com a Organização Mundial de Saúde, *mobile health (mHealth)* é definido como “a prática clínica ou de saúde pública apoiada por dispositivos móveis, como telefones celulares, dispositivos para o monitoramento de pacientes, assistentes digitais pessoais e outros dispositivos sem fio”¹. Por extensão à definição de cuidado centrado no paciente, pode-se dizer que tecnologias *mHealth* centrado no paciente são aquelas apoiadas por dispositivos móveis projetados para promover o cuidado centrado no paciente e, portanto, um dos seus principais usuários são os consumidores de saúde digital [BRUCE et al., 2020].

Ao mesmo tempo que os aplicativos *mHealth* trazem benefícios aos consumidores, eles podem trazer riscos, se eles forem imprecisos e não confiáveis, principalmente como vimos anteriormente esses usuários usam as informações para a tomada de decisão. Na revisão realizado por Akbar et. al [AKBAR; COIERA.; MAGRABI, 2020] foram encontrados os seguintes problemas na qualidade do conteúdo de alguns aplicativos de saúde:

¹ Tradução livre feito pelos autores

- a. Informação incorreta,
- b. Informação incompleta,
- c. variação no conteúdo entre aplicativos de domínios semelhantes,
- d. resultados incorretos, após cálculos e análises de respostas dos usuários,
- e. respostas inapropriadas às necessidades dos consumidores,

por exemplo, foram encontrados aplicativos que em caso de suspeita de suicídio não indicavam a busca de uma ajuda profissional nos momentos críticos.

Nesta mesma revisão foram encontrados estudos com falhas na funcionalidade do software:

- a. funcionalidades inadequadas de acordo com a tarefa, por exemplo, não permitir a entrada da dosagem em gramas ao invés de mililitros,
- b. falta de validação na entrada dos dados,
- c. processamento demorado que podem trazer risco à segurança do paciente, por exemplo, a demora no monitoramento de sinais vitais,
- d. resposta que trazem perigo à saúde,
- e. falha no alarme.

Além disso, foram encontrados vários aplicativos que não tinham o envolvimento de um especialista, ou em que o aplicativo não foi baseado em evidências clínicas e científicas e o mais preocupante deles é que não houve uma validação formais do aplicativo que pudessem trazer confiabilidade no seu conteúdo ou uso. Com isso fica evidente a necessidade de meios para que os profissionais de saúde e consumidores de saúde digital avaliem os aplicativos e assim possam escolher o aplicativo que mais seja confiável e adequado às suas necessidades [AKBAR; COIERA,; MAGRABI, 2020].

3.2 Avaliação de software

Um componente importante para assegurar a qualidade do software é uma avaliação de software. Incluindo o desenvolvimento de uma avaliação de qualidade de teste onde abrange a conformidade, estabilidade, adequação e eficácia [LU et al., 2022]. A qualidade de um software se refere à conformidade com os requisitos funcionais e de desempenho explicitamente declarados, a padrões de desenvolvimento claramente documentados e às características implícitas que são esperadas de todo software [MAXIM; PRESSMAN, 2021]. Os modelos de qualidade permitem avaliar softwares de acordo com diferentes aspectos e comumente representam as características desejáveis a um software em uma estrutura hierárquica [YANG, 2012].

Em outras palavras, a avaliação de software é um processo importante que visa analisar e medir a qualidade, eficácia e desempenho de um sistema ou aplicativo. Esta é uma prática ampla que visa garantir que o software satisfaça os requisitos estabelecidos e atinja os objetivos propostos. Existem várias dimensões e critérios envolvidos nesse processo, entre eles a usabilidade, confiabilidade, eficiência, manutenibilidade e portabilidade. A usabilidade, por exemplo, refere-se à facilidade de uso e compreensão do software pelo usuário final. Avaliações nessa dimensão consideram a interface do usuário, a intuitividade das operações e a experiência geral de interação. Já a confiabilidade está relacionada à capacidade do software de desempenhar suas funções de maneira consistente e livre de falhas, enquanto a eficiência diz respeito ao desempenho do software em termos de consumo de recursos e tempo de resposta. [ENTRINGER; FERREIRA, 2019].

A avaliação de software também abrange a manutenibilidade, que avalia a facilidade com que o software pode ser modificado ou corrigido, e a portabilidade, que se refere à capacidade do software de ser transferido entre diferentes ambientes. Além desses, outros critérios específicos podem ser considerados dependendo do contexto, como segurança, escalabilidade e interoperabilidade.

Existem diversas metodologias e ferramentas para a avaliação de software, desde testes manuais até abordagens mais automatizadas, como testes de unidade, integração e aceitação. A escolha da abordagem dependerá dos objetivos da avaliação, do tipo de software em questão e das características do projeto.

No campo da arquitetura de software, por exemplo, uma estrutura foi proposta para ajudar arquitetos de software iniciantes a avaliar os requisitos de qualidade de um sistema. Essa estrutura usa técnicas de processamento de imagem e aprendizado de máquina para inferir atributos e táticas de qualidade com base em padrões arquitetônicos [KAPUR et al., 2021].

Outro exemplo, é um método de avaliação da qualidade estática de software baseado no método AHP, que avalia objetiva e quantitativamente a qualidade do software e fornece orientação para a manutenção do software [MINGLIANG et al., 2021].

Porém, a avaliação de software, concentra-se especificamente na análise de programas de computador, aplicativos ou sistemas de software. Tem como principal fim analisar e determinar a qualidade, eficiência e adequação de um programa de software específico para cumprir seus objetivos.

Com algumas limitações, está mais restrita à análise das características e funcionalidades específicas do software, avaliando sua capacidade de realizar tarefas específicas. Por isso, faz uso de métodos mais específicos, como testes de software, revisões de código, avaliações de usabilidade e conformidade com requisitos.

3.3 Qualidade de software

Existem muitas definições de qualidade de software dispostas na literatura, sob diferentes pontos de vista de vários autores:

- A qualidade de um produto de software é medida com base em vários fatores para garantir que ele atenda às necessidades e expectativas do cliente, conforme discutido usando técnicas de sistema de inferência difusa no padrão ISO [BARZEGAR, 2021].
- A qualidade do software é determinada atendendo às necessidades do usuário por meio de processos e métricas bem definidos, como métricas de projeto, produto e processo, conforme discutido no artigo sobre melhoria da qualidade de software [KARUNA; DIVYA; MANGALA, 2018].
- Qualidade de software é a conformidade a requisitos funcionais e de desempenho que foram explicitamente declarados, a padrões de desenvolvimento claramente documentados, e a características implícitas que são esperadas de todo software desenvolvido por profissionais [MAXIM; PRESSMAN, 2021].

O conceito de qualidade de software engloba três fatores importantes, são eles os requisitos de software, os padrões e os requisitos implícitos. Diante disso a ISO/IEC 9126 afirma que qualidade é a totalidade de características e critérios de um produto ou serviço

que exercem suas habilidades para satisfazer às necessidades declaradas ou envolvidas. Baseado nessa fundamentação, a base para qualidade de software é descrita na FIGURA 3.1:

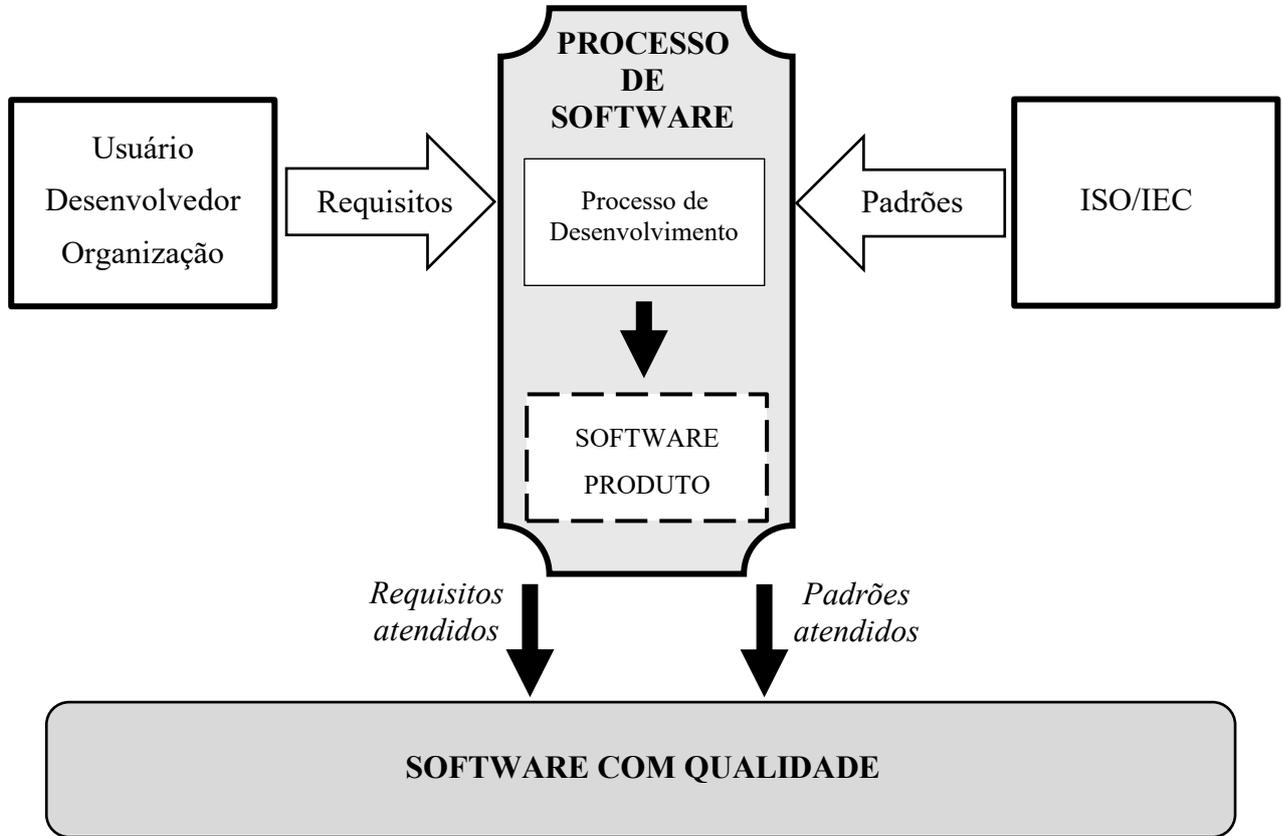


Figura 3.1 - Requisitos de software (Base da Qualidade)

Por meio da FIGURA 3.1, compreende-se que a qualidade de software é referente a atender plenamente os requisitos do cliente, ou seja, atingir e se possível superar as expectativas do cliente. Garvin [GARVIN, 2002] descreve as oito dimensões da qualidade de software na FIGURA 3.2:



Figura 3.2 - As oito dimensões da qualidade de Garvin

Por meio da figura acima, compreende-se que a qualidade de software é referente a atender plenamente os requisitos do cliente, ou seja, atingir e se possível superar as expectativas do cliente. Garvin [GARVIN, 2002] descreve as oito dimensões da qualidade de software na FIGURA 3.2:

- **Desempenho:** refere-se à capacidade do software de atender às expectativas dos usuários em termos de velocidade, precisão, eficiência e capacidade de resposta. Um software com bom desempenho é capaz de executar suas tarefas de forma rápida e eficiente, sem atrasos ou erros significativos.
- **Características:** são as funcionalidades e atributos essenciais do software que atendem às necessidades dos usuários. As características do software devem ser projetadas para proporcionar uma experiência satisfatória ao usuário, oferecendo as ferramentas e recursos necessários para realizar suas tarefas de forma eficaz.
- **Confiabilidade:** a probabilidade de o software funcionar sem falhas durante um período de tempo determinado concerne à confiabilidade. Um software confiável é aquele que pode ser utilizado de forma consistente e previsível, sem apresentar erros ou interrupções inesperadas que possam prejudicar a experiência do usuário.
- **Conformidade:** ao relacionar ao grau em que o software atende às normas, especificações e requisitos predefinidos, contempla a conformidade. Um software que está em conformidade com os padrões estabelecidos é aquele que cumpre todas as diretrizes e regulamentações relevantes, garantindo sua compatibilidade e interoperabilidade com outros sistemas e tecnologias.
- **Durabilidade:** é a capacidade do software de resistir ao desgaste e ao uso contínuo por um longo período de tempo. Um software durável é aquele que pode ser utilizado de forma consistente ao longo do tempo, sem perder sua eficácia ou desempenho devido ao uso prolongado ou condições adversas.
- **Facilidade:** relaciona-se à facilidade com que os usuários podem aprender, operar e entender o software. Um software fácil de usar é aquele que possui uma interface intuitiva e amigável, permitindo que os usuários realizem suas tarefas sem a necessidade de treinamento extensivo ou assistência técnica.
- **Estética:** são os aspectos visuais do software que influenciam a percepção do usuário em termos de atratividade e prazer de uso. Um software esteticamente atraente é aquele que possui um design visualmente agradável, com layouts, cores e elementos gráficos que tornam a experiência do usuário mais agradável e envolvente.
- **Qualidade percebida:** pertence a avaliação subjetiva da qualidade do software feita pelo usuário, considerando todos os outros fatores mencionados. A qualidade percebida de um software é influenciada pela experiência do usuário, suas expectativas e percepções sobre a eficácia, confiabilidade, facilidade de uso e outros aspectos do software.

A partir do conteúdo exposto na FIGURA 3.1, compreende-se que a qualidade de software pode ser definida como a relação entre os requisitos e os resultados alcançados, ou seja, pode ser considerada como a diferença entre o que se espera, e o que se consegue.

Por fim, conforme Roos, Sartori e Godoy [ROOS, SARTORI E GODOY, 2009], utilizando o modelo de Kano, define a qualidade de software em três níveis: qualidade por excelência, qualidade unidimensional e qualidade esperada. Onde:

- **Qualidade por excelência** é definida como qualquer característica do software que exceda o nível de desempenho esperado.
- **Qualidade unidimensional** é definida como qualquer característica do software desejada por usuários específicos, porém não inteiramente necessária para todos as categorias de usuários.
- **Qualidade esperada** é definida como qualquer característica necessária ao funcionamento normal do produto.

Diante de tudo isso, é possível verificar que é imprescindível o desenvolvimento de softwares de qualidade que sejam capazes de suprir e atender a todos os requisitos exigidos pelo cliente, e como meio de reger a qualidade faz-se necessário o uso de normas que estabelecem regras e critérios para assegurar que os produtos de software estejam adequados para o devido uso.

3.3.1 Conceitos Fundamentais em Avaliação de Software

Conceitos fundamentais na avaliação de software abrangem vários aspectos, como obter solicitações de fontes externas, analisar saídas de log, testar com diversos dados de entrada, identificar modificações na estrutura de dados e avaliar vulnerabilidades de softwares [MINOV, 2012] [YAMAKOSHI; NISHIYAMA, 2018]. A avaliação de desempenho envolve pontos de vista, métricas, caracterização da carga de trabalho e benchmarking [OBAIDAT; BOUDRIGA, 2010].

Além disso, os modelos de estimativa de software desempenham um papel crucial no gerenciamento de software, enfatizando o design e a avaliação de modelos de estimativa [ABRAN, 2015]. O teste, uma parte vital da avaliação de software, visa garantir que os sistemas atendam aos requisitos especificados por meio de execução sistemática e identificação de erros [YADAV, 2015]. Esses conceitos combinados formam a base da avaliação de software, abrangendo testes, métricas de desempenho, análise de entradas externas e avaliação de vulnerabilidades.

3.3.2 Métodos e Critérios de Avaliação

A avaliação de produto de software com base em normas de qualidade tem sido uma das formas empregadas por organizações que produzem ou adquirem software para aferirem a qualidade de seus produtos. Assim, para que a avaliação seja mais efetiva, é importante a utilização de modelos de qualidade que permitam estabelecer e avaliar requisitos de qualidade, e que o processo de avaliação seja bem definido e estruturado.

Assim, para que esses requisitos sejam atendidos e a qualidade atingida, existem as normas ISO/IEC 9126, ISO/IEC 14598 e ISO/IEC 25000 que trabalha com especificações técnicas, que estabelecem regras e critérios para assegurar que os produtos de software estejam adequados para o devido uso, desta forma essas normas tornam-se essenciais, pois elas orientam para o devido desenvolvimento e avaliação dos produtos de software com o objetivo de atender a expectativa e qualidade exigida pelo cliente [ISO 9126, 2001] [ISO 14598, 1998] [ISO 25000, 2014].

Algumas normas foram criadas para padronizar internacionalmente as características de implementação e avaliação da qualidade de software.

Para melhor entendimento, segue nos próximos itens especificações sobre as normas ISO/IEC 9126, ISO/IEC 14598 e ISO/IEC 25000 que trabalham com modelo de avaliação de software e métricas para aperfeiçoamento.

3.3.2.1 ISO/IEC 9126, define um modelo de qualidade para avaliação de produtos de software

A norma ISO/IEC 9126 recomenda que os produtos de software sejam avaliados usando um modelo de qualidade, que tem a finalidade de aprimorar o produto, para que dessa forma seja possível cumprir os requisitos para atingir a expectativa do cliente [ISO 9126, 2001]. A série de normas ISO/IEC 9126 descreve um modelo de qualidade para produtos de software categorizando a qualidade hierarquicamente em um conjunto de características e subcaracterísticas. Esta série também propõe métricas que pode ser utilizada durante a avaliação dos produtos de software (medição, pontuação e julgamento dos produtos de software).

A ISO/IEC 9126 é dividida em quatro partes (FIGURA 3.3): (9126-1) modelo de qualidade; (9126-2) métricas externas; (9126-3) métricas internas e, (9126-4) métricas de qualidade de uso.

- **ISO/IEC 9126-1** é definido um modelo de qualidade para produtos de software.
- **ISO/IEC 9126-2** é proposto métricas externa para medir os atributos das características de qualidade definidas na ISO/IEC 9126-1.
- **ISO/IEC 9126-3** apresenta métricas internas para medir os atributos das características de qualidade definidas na ISO/IEC 9126-1.
- **ISO/IEC 9126-4** descreve sobre métricas de qualidade em uso para medir os atributos das características de qualidade definidas na ISO/IEC 9126-1.

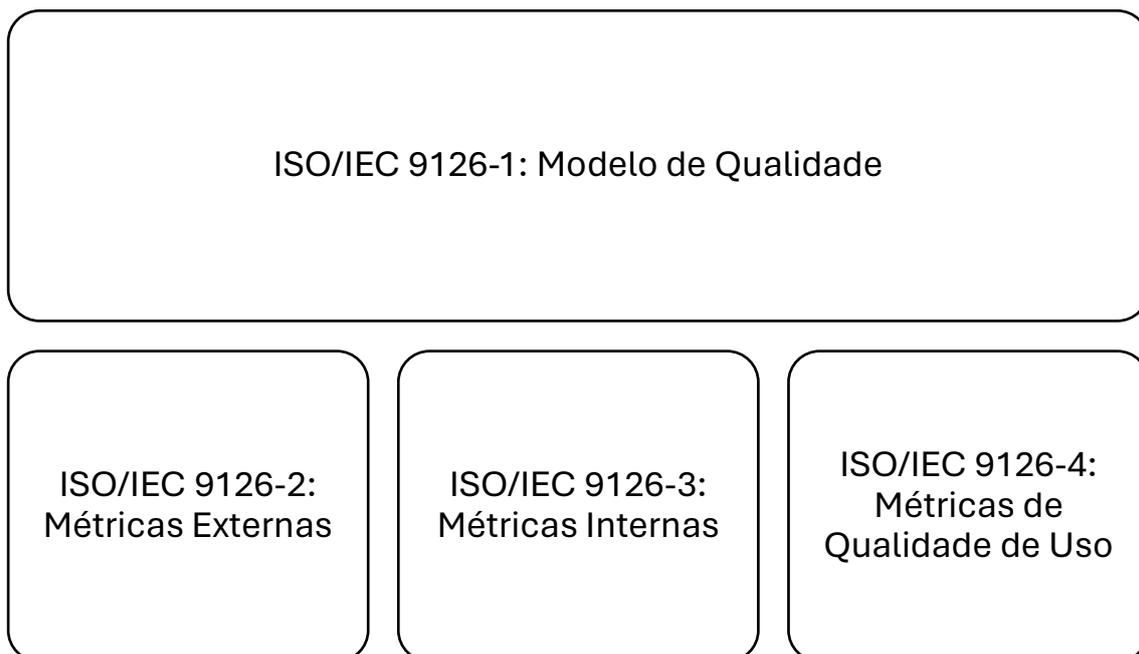


Figura 3.3 - Estrutura da série de normas ISO/IEC 9126

Nas ISO/IEC 9126-2 e 9126-3 apresenta um modelo de qualidade externa e interna que separa os atributos de qualidade de software em seis características: (i) funcionalidade; (ii) confiabilidade; (iii) usabilidade; (iv) eficiência; (v) manutenibilidade e (vi) portabilidade (FIGURA 3.4).

Essas características são subdivididas em subcaracterísticas, as quais podem ser mensuradas por métricas internas e externas, conforme exposto na FIGURA 3.4:

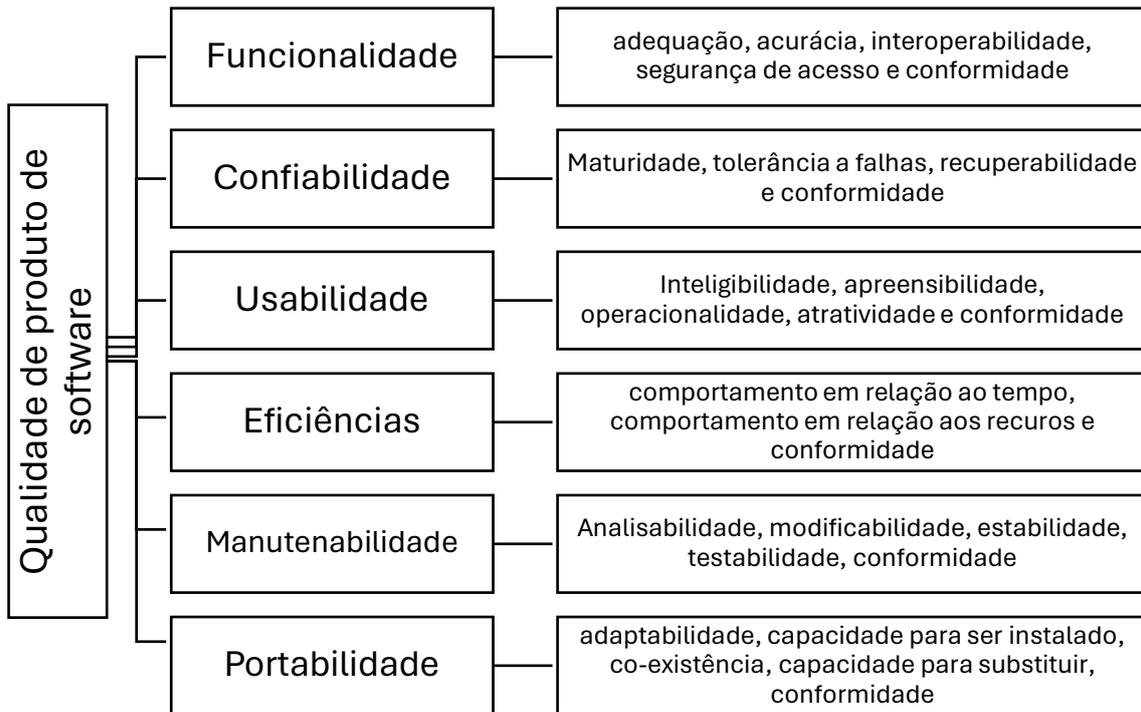


Figura 3.4 - ISO/IEC 9126-2 e 9126-3: Características e Subcaracterísticas

A norma ISO/IEC 9126 – 4 apresenta um modelo para qualidade. O modelo de qualidade em uso categoriza os atributos de qualidade de software em quatro características (Eficácia, Produtividade, Segurança e Satisfação). A qualidade em uso é a visão da qualidade sob a perspectiva do usuário. A obtenção de qualidade em uso é dependente da obtenção da necessária qualidade externa, a qual, por sua vez, é dependente da obtenção da necessária qualidade interna. Normalmente, são necessárias medidas em todos os três níveis, pois atender aos critérios para medidas internas em geral não é suficiente para garantir o atendimento aos critérios para medidas externas, e atender aos critérios para medidas externas de subcaracterísticas em geral não é suficiente para garantir o atendimento aos critérios para qualidade em uso [ISO 9126, 2001].

A ISO/IEC 9126 define as características do modelo da qualidade em uso da seguinte maneira:

- **Eficácia:** capacidade do produto de software de permitir que usuários atinjam metas especificadas com acurácia e completude, em um contexto de uso especificado.
- **Produtividade:** capacidade do produto de software de permitir que seus usuários empreguem quantidade apropriada de recursos em relação à eficácia obtida, em um contexto de uso especificado.

- **Segurança:** capacidade do produto de software de apresentar níveis aceitáveis de riscos de danos a pessoas, negócios, software, propriedades ou ao ambiente, em um contexto de uso especificado.
- **Satisfação:** capacidade do produto de software de satisfazer usuários, em um contexto de uso especificado

Em síntese, a ISO/IEC 9126 é uma norma composta por um conjunto de características que devem ser verificadas em um software para que ele seja considerado um “software de qualidade”, bem como, métricas usadas na sua avaliação (medição, pontuação e julgamento) dos softwares.

3.3.2.2 ISO/IEC 14598-1 - Tecnologia de informação – Avaliação de produto de software

O processo de avaliação da ISO/IEC 14598 é baseado na norma ISO/IEC 9126, a qual estabelece as métricas de qualidade de software e pode ser usado tanto para avaliar produtos prontos como produtos em desenvolvimento.

A avaliação de produtos de software deve ser objetiva, ou seja, baseada em observação e não em opinião. Também deve ser reprodutiva, de forma que avaliações do mesmo produto, para a mesma especificação de avaliação, executadas por diferentes avaliadores produzam resultados aceitos como idênticos e repetíveis. Para isso, um processo de avaliação deve ser definido. Este processo deve seguir, basicamente, cinco passos: análise dos requisitos de avaliação, especificação da avaliação, projeto e planejamento da avaliação, execução da avaliação e documentação dos resultados. A série de normas ISO/IEC 14598 descreve um processo para avaliação de produtos de software, que consiste em quatro passos: análise especificação, projeto e execução (TABELA 3.1). O padrão definido distingue três perspectivas de avaliação: desenvolvedor, adquirente e avaliador [ISO 14598, 1998].

Tabela 3.1 - Definição do processo de avaliação segundo a ISO/IEC 14598.

	Análise	Especificação	Projeto	Execução
Processo para desenvolvedores	Definição de requisitos de qualidade e análise de sua exequibilidade	Quantificação dos requisitos de qualidade	Planejamento da avaliação durante o desenvolvimento	Monitoramento da qualidade e controle durante o desenvolvimento
Processo para adquirentes	Estabelecimento do propósito e escopo da avaliação	Definição de métricas externas e medições correspondentes a serem realizadas	Planejar, programar e documentar a avaliação	A avaliação deveria ser realizada, documentada e analisada
Processo para avaliadores	Descrição dos objetivos da avaliação	Definição do escopo da avaliação e das medições	Documentação dos processos a serem usados pelo avaliador	Obtenção dos resultados a partir da realização de ações de medição e verificação do produto

O processo de avaliação da ISO/IEC 14598 é realizado em seis etapas, sendo a primeira uma abordagem geral de como é realizada a avaliação, seguida do planejamento

e intercursos, dos papéis dos desenvolvedores, dos adquirentes e dos avaliadores, e por último da documentação dos módulos de avaliação.

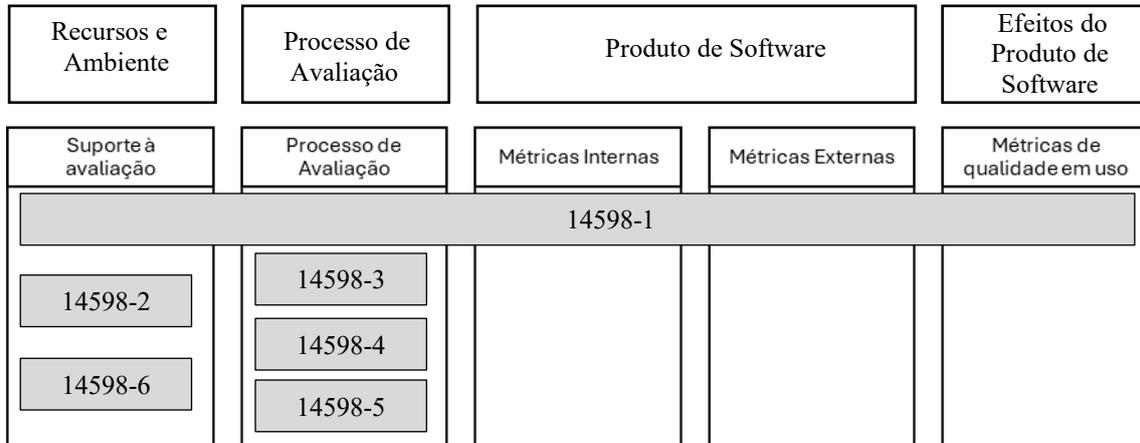


Figura 3.5 - Estrutura ISO/IEC 14598

A série de normas ISO/IEC 14598 é dividida em seis partes (FIGURA 3.5 e FIGURA 3.6):

- **ISO/IEC 14598-1 - Visão geral:** fornece uma visão geral do processo de avaliação da qualidade dos produtos de software e define toda a estrutura de funcionamento da série de normas ISO/IEC 14598.
- **ISO/IEC 14598-2 - Planejamento e gestão:** refere-se ao planejamento e gestão do processo de avaliação apresentando requisitos, recomendações e orientações para uma função de suporte ao processo
- **ISO/IEC 14598-3 - Processo para desenvolvedores:** define o processo para desenvolvedores. Destina-se ao uso durante o processo de desenvolvimento e manutenção de software
- **ISO/IEC 14598-4 - Processo para adquirentes:** define o processo para adquirentes, estabelecendo um processo sistemático para avaliação de: produtos de software tipo pacote (com equivalência a NBR ISO/IEC 12119), produtos de software sob encomenda, ou ainda modificações em produtos já existentes
- **ISO/IEC 14598-5 - Processo para avaliadores:** define o processo para avaliadores, fornecendo orientações para a implementação prática de avaliação de produtos de software (quando diversas partes necessitam entender, aceitar e confiar em resultados da avaliação)
- **ISO/IEC 14598-6 - Documentação de módulos para avaliação:** fornece orientação para documentação de módulos de avaliação. Estes módulos contêm a especificação do modelo de qualidade, as informações e dados relativos à aplicação prevista do modelo e informações sobre a real aplicação do modelo

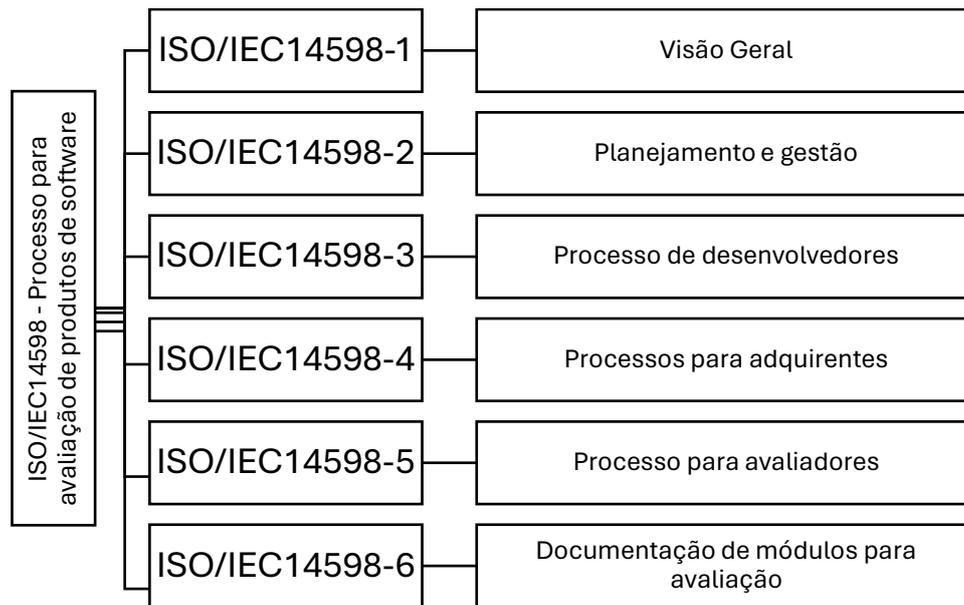


Figura 3.6 - Estrutura da série de normas ISO/IEC14598

Diante dessas séries da norma, compreende-se que para avaliar a qualidade do software, primeiramente devem ser estabelecidos os requisitos da avaliação, para então poder especificar, projetar e executar a avaliação, como exposto na FIGURA 3.7:

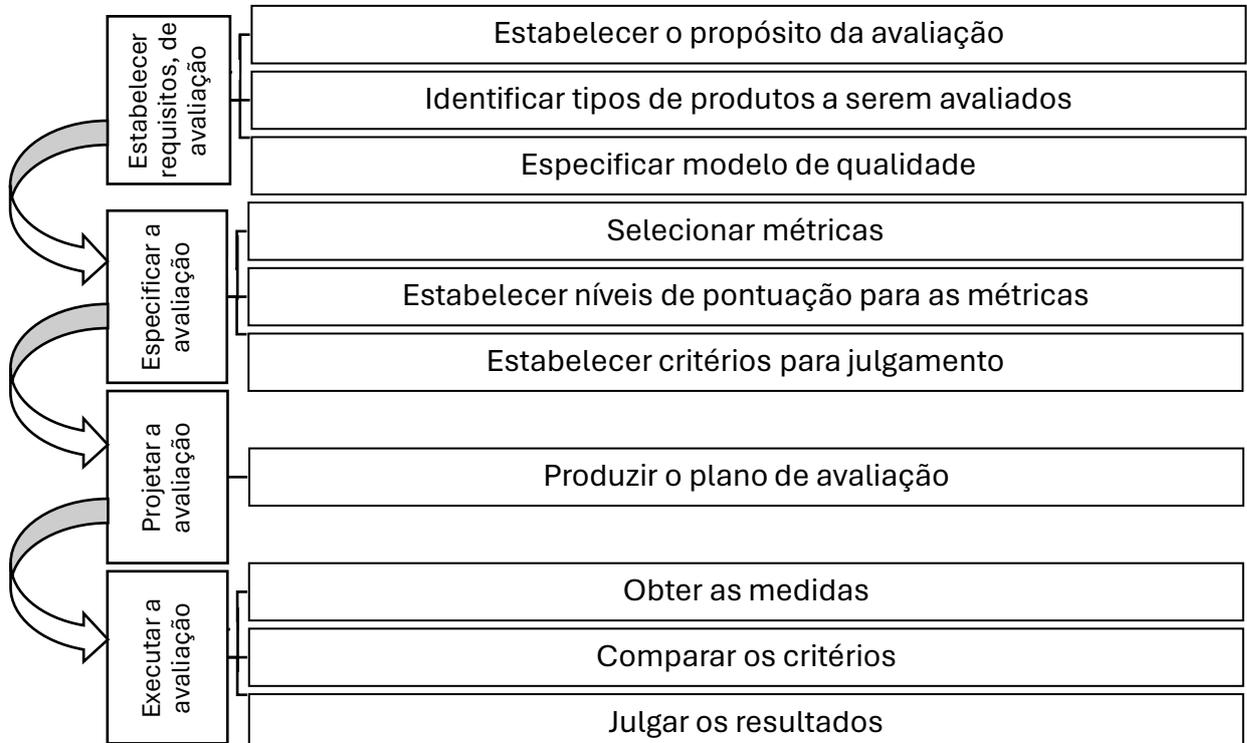


Figura 3.7 - Processo de avaliação segundo a ISO/IEC 14598-1

Compreende-se que a intenção da norma ISO 14598 é estabelecer uma padronização nos módulos de avaliação para que eles possam ser reusáveis. Criam-se assim, bibliotecas desses módulos, fornecendo também guias para documentação dos módulos e suporte ao seu desenvolvimento. O modelo com que as informações são manipuladas ou quais informações são necessárias na realização da avaliação, ilustram o processo de documentação dos módulos de avaliação.

3.3.2.3 ISO/IEC 25000 - SQuaRE (Requisitos e Avaliação de Qualidade de Sistemas e Software)

Seguindo a ISO 25000, que unifica as normas ISO/IEC 9126 e ISO/IEC 14598, desenvolver os processos principais: especificação de requisitos e a avaliação da qualidade.

O ISO 25000 é um padrão bem conhecido no campo da qualidade de software, que inclui o padrão ISO/IEC 25030 que ajuda a especificar os requisitos de qualidade de software. Os padrões ISO visam garantir que o desenvolvimento de software atenda a critérios de alta qualidade e produza aplicativos eficazes e versáteis. Esses padrões fornecem uma estrutura para entender, medir e prever a qualidade do software, facilitando para as organizações o desenvolvimento de produtos de software de qualidade. O trabalho da ISO na padronização da qualidade de software ajuda as organizações a reduzirem os custos de manutenção e melhorar a qualidade geral de seus produtos de software. Seguindo os padrões ISO, as organizações podem estabelecer uma base para a qualidade desejada e se beneficiar de um conceito contínuo e hierárquico de qualidade em todo o processo de desenvolvimento de software [KUI; ALI; SURYN, 2016] [ZHANG et al., 2018] [BOUGROUN et al., 2020] [BARZEGAR, 2021].

A série de normas ISO/IEC 25000:2014 (Software Quality Requirements and Evaluation, SQuaRE), dividida em cinco partes, foi desenvolvida para fornecer suporte à definição de requisitos e ao processo de avaliação da qualidade de softwares. O modelo de qualidade apresentado na ISO/IEC 25010:2014 aborda duas dimensões da qualidade de software.

A primeira delas é a qualidade em uso, que especifica características de qualidade relacionadas à interação humana com o sistema. A qualidade em uso se refere à capacidade de atender a requisitos para atingir metas específicas com produtividade, efetividade, segurança e satisfação do usuário, em um contexto de uso especificado (ISO 25000, 2014). A segunda dimensão proposta pela ISO/IEC 25010 é a qualidade do produto, que define um conjunto de oito características da qualidade relacionadas à atributos internos e externos do software. A adequação funcional é a capacidade do software de prover funções que atendam às necessidades implícitas e explícitas quando usado em condições especificadas. A eficiência de desempenho se refere ao desempenho em relação à quantidade de recursos utilizados. A confiança é a capacidade de manter um nível de desempenho especificado em um determinado período de tempo. A usabilidade é a capacidade de ser usado para atingir metas específicas com efetividade, produtividade e satisfação do usuário. A manutenibilidade é a capacidade de ser modificado visando à melhoria, correção ou adaptação a mudanças no ambiente ou nos requisitos. A portabilidade é a capacidade de ser transferido de um hardware, sistema operacional ou ambiente de uso para outro. A compatibilidade é a capacidade de trocar informações com outros softwares e desempenhar as funções que lhe forem requeridas enquanto

compartilha recursos de hardware e software. A segurança diz respeito à capacidade do software de proteger informações para que pessoas ou outros softwares tenham o nível de acesso apropriado aos seus níveis de permissão [ISO 25000, 2014]. Conforme descreve a Tabela 3.2, o modelo de qualidade apresentado na ISO/IEC 25010 desdobra as características da qualidade em um conjunto de subcaracterísticas relacionadas.

Tabela 3.2 - Características e subcaracterísticas da qualidade de software.

CARACTERÍSTICA	SUBCARACTERÍSTICA	DEFINIÇÃO DA SUBCARACTERÍSTICA
Adequação funcional	Integridade funcional	Capacidade de atender às tarefas e aos objetivos específicos do usuário a que foi destinado
	Correção funcional	Capacidade de fornecer resultados corretos e com precisão
	Adequação funcional	Capacidade de realizar tarefas e certos objetivos de maneira fácil
Eficiência de desempenho	Comportamento em relação ao tempo	Capacidade de fornecer tempos de resposta e processamento apropriados quando o software executa suas funções
	Utilização de recursos	Grau com que os tipos e as quantidades de recursos usados atendam aos requisitos
	Capacidade	Grau com que a capacidade máxima de parâmetros do software atende aos requisitos
Usabilidade	Reconhecibilidade	Grau com que o usuário é capaz de reconhecer se o produto é adequado às suas necessidades
	Apreensibilidade	Capacidade que o software possui de ser usado para alcançar objetivos específicos de aprendizagem, de forma eficiente, eficaz e sem riscos, garantindo que o usuário se sinta satisfeito no contexto em questão
	Operacionalidade	Capacidade do software de permitir ao usuário operá-lo e controlá-lo de forma fácil
	Proteção de erro	Capacidade de proteger os usuários de cometer falhas
	Estética da interface	Capacidade de possuir uma interface que seja satisfatória ao usuário
	Acessibilidade	Capacidade de ser usado por usuários com diferentes características e habilidades para alcançar objetivos especificados

CARACTERÍSTICA	SUBCARACTERÍSTICA	DEFINIÇÃO DA SUBCARACTERÍSTICA
Compatibilidade	Coexistência	Capacidade de compartilhar recursos com outro software sem causar impactos sobre qualquer outro produto
	Interoperabilidade	Capacidade de interagir com um ou mais sistemas especificados
Confiança	Maturidade	Capacidade de atender às necessidades de confiabilidade quando operado em condições normais
	Disponibilidade	Capacidade de ser operacional e acessível quando requerido para uso
	Tolerância a falhas	Capacidade de garantir um nível de desempenho especificado em caso de ocorrência de falhas de software ou hardware
	Capacidade de recuperação	Capacidade de restabelecer seu nível de desempenho especificado e recuperar os dados diretamente afetados no caso de uma falha
Segurança	Confidencialidade	Capacidade de permitir acesso de dados somente a usuários autorizados
	Integridade	Capacidade de bloquear acesso e modificações de usuários não autorizados
	Não repúdio	Capacidade de comprovar ações, eventos, alterações e envio de informações para que não possam ser repudiados futuramente
	Responsabilidade	Capacidade de rastrear as ações de entidades específicas
	Autenticidade	Capacidade de comprovar a identidade de um sujeito ou recurso caso seja requerido
Manutenibilidade	Modularidade	Capacidade de alterar elementos do software com impacto mínimo
	Reutilização	Capacidade que os componentes do software possuem de serem utilizados por outros sistemas existentes ou em construção
	Analisabilidade	Capacidade de avaliar o impacto de uma mudança em uma ou mais partes de um sistema, diagnosticar partes do sistema em que pode haver falhas e identificar componentes a serem modificados
	Modificabilidade	Capacidade de permitir que uma modificação seja implementada sem causar defeitos no produto existente

CARACTERÍSTICA	SUBCARACTERÍSTICA	DEFINIÇÃO DA SUBCARACTERÍSTICA
	Testabilidade	Capacidade de estabelecer critérios de teste para um sistema que foi modificado e de determinar se esses critérios foram cumpridos de forma eficaz e eficiente
Portabilidade	Adaptabilidade	Capacidade de ser adaptado para ambientes de operação especificados, sem a necessidade de aplicação de outras ações ou meios além daqueles fornecidos para essa finalidade pelo software considerado
	Instalabilidade	Capacidade de ser corretamente instalado e/ou desinstalado em um ambiente especificado
	Substituibilidade	Capacidade de substituir outro software no mesmo ambiente e para o mesmo fim

Embora a ISO/IEC 25010 sugira um conjunto de 31 subcaracterísticas da qualidade, a realização da avaliação de um ou mais softwares usando todas essas subcaracterísticas pode se tornar inviável devido ao tempo consumido. Dessa forma, essas características e subcaracterísticas devem servir de referência para a criação de modelos de avaliação adequados ao tempo, aos recursos humanos e às informações disponíveis [JUNG, 2007].

Na literatura podem ser encontrados diversos modelos de avaliação da qualidade para tipos de software específicos. Silva et al. [2015] desenvolveram um modelo para avaliar a qualidade de um sistema acadêmico com base nos critérios facilidade de uso, comportamento em relação a falhas, eficiência e satisfação das necessidades. Lima Junior e Martimiano [2010] propuseram um modelo de apoio à avaliação da qualidade de softwares PMIS com base na norma ISO/IEC 9126. Uma limitação desse modelo é a impossibilidade de avaliar os níveis de importância relativa (ou pesos) das características e subcaracterísticas da qualidade. Como os requisitos dos usuários variam de acordo com o contexto de aplicação do software, é desejável atribuir pesos diferentes para as características da qualidade [YANG, 2012]. Esse processo de ponderação pode ser apoiado por métodos de tomada de decisão multicritério [TRENDOWICZ; KOPCZYNSKA, 2014].

3.4 Avaliação de Tecnologia em Saúde Digital

A quantidade de medicamentos, dispositivos, procedimentos e outras tecnologias em saúde é vasta, e há um contínuo desenvolvimento de novas tecnologias para a atenção à saúde. Esse desenvolvimento de novas tecnologias contribui para a melhoria da qualidade de vida, e pressionam os sistemas de saúde a incorporá-las, que contam com recursos limitados. A inovação, regulação e garantia de acesso de tecnologias em saúde são fatores importantes para a sustentabilidade dos sistemas de saúde.

Neste cenário, torna-se necessário que haja procedimentos para nortear o desenvolvimento e adoção dessas novas tecnologias de forma adequada e segura nos sistemas de saúde, garantindo efetividade, segurança e eficácia, ao mesmo tempo que

aquelas tecnologias que não são tão eficazes como outras, possam ser substituídas por estas.

Segundo a Portaria Nº 2.510/GM de 19 de dezembro de 2005 [GABINETE DO MINISTRO, 2005], considera-se tecnologias em saúde: medicamentos, materiais, equipamentos e procedimentos, sistemas organizacionais, educacionais, de informações e de suporte, e programas e protocolos assistenciais, por meio dos quais a atenção e os cuidados com a saúde são prestados à população.

No tocante à inovação, o desenvolvimento de tecnologias da informação e comunicação ajudou a cunhar o termo saúde digital. A saúde digital se refere ao uso de tecnologia para diagnosticar, monitorar, tratar e prevenir doenças. Abrange vários avanços, como saúde móvel, registros eletrônicos de saúde, dispositivos vestíveis, telessaúde, inteligência artificial e robótica [BUCHNER, 2022].

O conceito de saúde digital está transformando os modelos tradicionais de prestação de cuidados de saúde, levando a uma transformação cultural no campo [MESKÓ et al., 2017]. No entanto, a implementação da saúde digital enfrenta desafios como regulamentações rígidas, resistência à mudança e a necessidade de considerações culturais e de fatores humanos [SHAFFER; PREZIOSI, 2019]. Além disso, há riscos associados à saúde digital, incluindo ameaças cibernéticas, perda de privacidade e disseminação de desinformação médica [SIMON LIN LINWOOD, 2022].

No geral, a saúde digital tem o potencial de revolucionar a saúde, mas uma consideração cuidadosa de seus riscos e benefícios é fundamental. A afirmação destaca a transformação potencialmente revolucionária proporcionada pela saúde digital no cenário da saúde. A expressão "saúde digital" refere-se à aplicação de tecnologias digitais, como dispositivos móveis, softwares e plataformas online, para melhorar a prestação de serviços de saúde. Embora essa revolução prometa avanços significativos, é importante abordá-la com uma consideração cuidadosa dos riscos e benefícios envolvidos. Isso implica avaliar minuciosamente as implicações éticas, privacidade dos dados, segurança cibernética e equidade no acesso, garantindo que o impacto positivo na saúde seja maximizado, ao mesmo tempo em que se minimizam possíveis desafios e ameaças. A busca por inovações na saúde digital deve, portanto, ser guiada por uma abordagem equilibrada e responsável para garantir seu pleno potencial em benefício da sociedade.

A saúde digital está relacionada à avaliação de várias maneiras. Em primeiro lugar, tecnologias digitais de saúde e dispositivos de biotelemetria estão sendo usados para prevenir, detectar, diagnosticar, monitorar e prever a evolução de várias doenças, sem exigir procedimentos invasivos ou interação face a face com o pessoal médico [BUCHNER, 2022].

Em segundo lugar, métodos e sistemas de avaliação de saúde digital discretos foram desenvolvidos, os quais integram características físicas, hábitos de vida e condições médicas prevalentes com atividades físicas monitoradas e medidas fisiológicas para avaliar a saúde de indivíduos de alto risco.

Em terceiro lugar, o impacto da digitalização dos serviços de saúde precisa ser avaliado com base no alcance das metas do sistema de saúde, incluindo qualidade, acessibilidade, eficiência e equidade [EUROPEAN COMMISSION. DIRECTORATE GENERAL FOR HEALTH AND FOOD SAFETY, 2019].

Além disso, a saúde digital tem beneficiado muito o sistema público de saúde, e a acessibilidade e a confiança dos pacientes nas informações da Internet influenciam suas decisões e avaliações dos profissionais de saúde [VUONG et al., 2022].

Finalmente, a Avaliação de Tecnologia de Saúde (ATS) de aplicativos móveis de saúde, incluindo aplicativos digitais de saúde, envolve uma avaliação rápida de sua eficácia e potencial de economia de custos para informar os tomadores de decisão no sistema de saúde [LESQUELEN et al., 2017].

Esta subseção apresenta os principais aspectos relacionados à avaliação de tecnologia em saúde digital.

3.4.1 Definição de Tecnologia em Saúde

O termo Tecnologias em Saúde abrange produtos, procedimentos e insumos utilizados nos serviços de saúde, e abrange medicamentos, vacinas, testes diagnósticos, equipamentos e dispositivos, modelos de organização de serviços de saúde, bem como os sistemas de apoio.

Segundo a Organização Mundial da Saúde, o termo Tecnologias de Saúde se aplica a próteses, diagnósticos in vitro, equipamentos de imagem e soluções de e-saúde utilizados para diagnosticar, monitorar, avaliar, prevenir e indicar o tratamento aos pacientes que sofrem de uma vasta gama de doenças. Reconhecendo que há muitas interpretações para os termos que tratam do setor, são apresentadas abaixo as definições estabelecidas pela Organização Mundial de Saúde (OMS) e pela Global Harmonization Task Force (GHTF) [SAÚDE 4.0, 2015].

- **Tecnologias médicas ou Dispositivos Médicos no sentido Amplo (DMAs):** envolvem a aplicação de conhecimento organizado e de práticas médicas sob a forma de aparelhos, procedimentos e sistemas desenvolvidos para identificar, prevenir e solucionar um problema de saúde e melhorar a qualidade de vida. Os DMAs incluem uma ampla variedade de produtos, indo desde os mais simples, como luvas, linhas de sutura, adesivos, camas hospitalares e lentes, até aplicativos para smartphones, implantes cardíacos, monitores de glicose e scanners de ressonância magnética. O prolongamento da vida mais ativa e independente, que vemos nos dias de hoje, é em larga medida o resultado das inovações na área da tecnologia médica. Abrangem qualquer instrumento, máquina, aparelho, implante, reagente para uso in vitro, software, material ou outro artigo similar ou relacionado, destinado pelo fabricante a ser utilizado, isoladamente ou em combinação pelos seres humanos, para um ou mais dos efeitos médicos específicos de:
 - diagnóstico, prevenção, controle, tratamento ou atenuação de uma doença;
 - diagnóstico, controle, tratamento, atenuação ou compensação de uma lesão;
 - estudo, substituição, modificação ou apoio da anatomia ou de um processo fisiológico;
 - apoiar ou sustentar a vida;
 - controle da concepção;
 - desinfecção de dispositivos médicos;
 - fornecimento de informações por meio de exame in vitro de amostras derivadas do corpo humano.

- **Dispositivos médicos no sentido restrito:** são artigos, instrumentos, aparatos ou máquinas utilizados na prevenção, diagnóstico ou tratamento de sintoma ou doença, com o objetivo de detectar, medir, restaurar, corrigir ou modificar a estrutura ou função do corpo para algum propósito de saúde. Tipicamente, o propósito de um dispositivo médico não é alcançado por meios farmacológicos, imunológicos ou metabólicos.
- **Equipamentos médicos:** são aparelhos médicos que requerem calibração, manutenção, reparo e treinamento de quem o utiliza. O equipamento médico é usado com o propósito específico de diagnosticar e tratar um problema de saúde. Pode ser usado sozinho ou em combinação com acessórios e produtos para consumo, e, dessa definição, são excluídos os produtos para a saúde implantáveis, descartáveis ou de uso único.
- **Diagnóstico in vitro:** que trata-se de qualquer dispositivo médico que consista num reagente, instrumento, aparelho ou sistema usado, isoladamente ou combinado, a ser utilizado in vitro no exame de amostras provenientes do corpo humano, de forma a obter informações sobre estados fisiológicos, de saúde, de doença, ou de anomalia congênita.
- **E-Saúde:** abrange as áreas e produtos para a saúde que se utilizam das tecnologias de informação e comunicação para diagnóstico, atendimento, automonitoramento, educação e prestação de serviços remotos em saúde. A informatização na saúde, tanto no controle de custos e na organização dos procedimentos, como na aferição da qualidade do atendimento ao paciente, configura-se como objeto de grande interesse em face dos significativos investimentos privados e públicos que requer e do seu potencial de bons resultados.

Na literatura são encontradas algumas formas de classificação de tecnologias de saúde. Em [LIAROPOULOS, 1997] é proposto uma hierarquia de tecnologias de saúde.

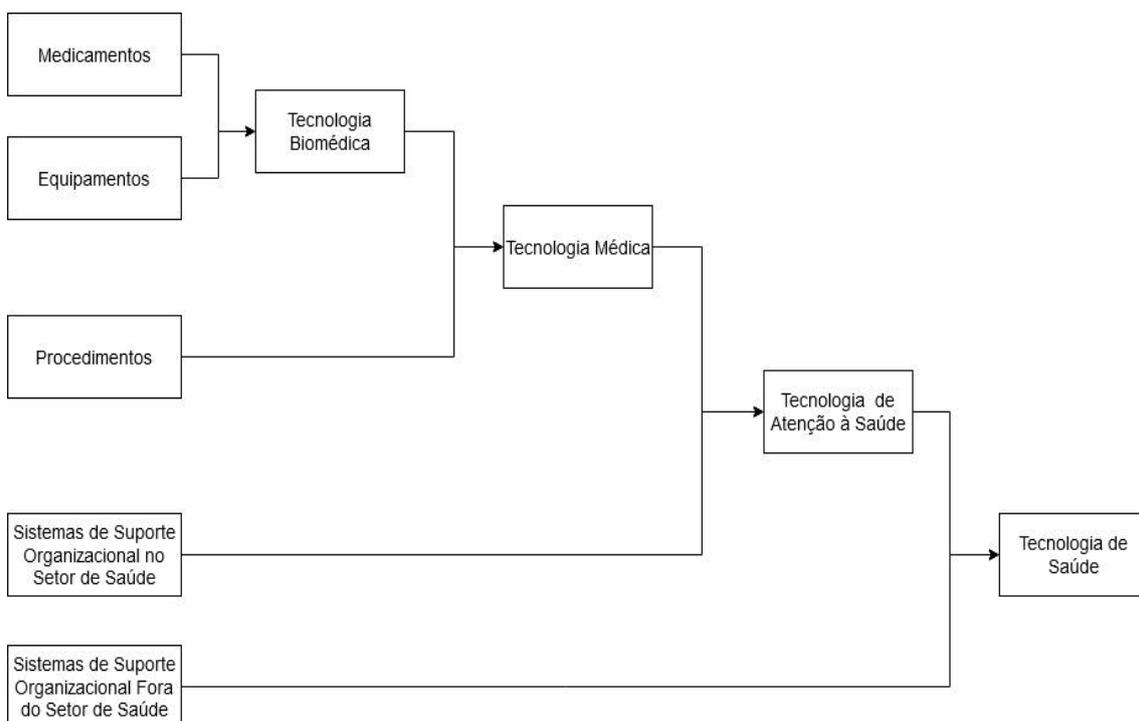


Figura 3.8 - Hierarquia de tecnologias de saúde

Medicamentos e equipamentos que interagem diretamente com pacientes são designados como tecnologias biomédicas. Os procedimentos, como anamnese, técnicas cirúrgicas, normas técnicas de uso de tecnologias biomédicas, são designadas, juntamente com as tecnologias biomédicas, tecnologias médicas. O uso das tecnologias médicas está inserido dentro de uma estrutura técnica e administrativa, contando também com sistemas de informação, denominados como Sistemas de Suporte Organizacional no Setor de Saúde. Os Sistemas de Suporte Organizacional no Setor de Saúde e as Tecnologias Médicas são designadas Tecnologias de Atenção à Saúde.

Por fim, há componentes organizacionais e de apoio que atuam fora do setor de saúde e que de alguma forma influenciam a saúde como, por exemplo, saneamento, controle ambiental e leis trabalhistas.

O uma outra classificação é proposta em [GOODMAN, 1998], de acordo com a natureza do material, o propósito, e o estágio de difusão:

- Medicamentos:
 - Equipamentos e suprimentos como ventiladores, marcapassos cardíacos, luvas cirúrgicas, kits de diagnóstico, etc;
 - Procedimentos médicos e cirúrgicos;
 - Sistemas de suporte: bancos de sangue, sistemas de prontuário eletrônico, etc.;
 - Sistemas gerenciais e organizacionais: sistema de informação, sistema de garantia de qualidade, etc.
- Quanto ao propósito:
 - Prevenção: visa proteger os indivíduos contra uma doença ou limitar a extensão de uma seqüela (exemplo: imunização, controle de infecção hospitalar, etc.);
 - Triagem: visa detectar a doença, anormalidade, ou fatores de risco em pessoas assintomáticas (mamografia, exame de Papanicolau);
 - Diagnóstico: visa identificar a causa e natureza ou extensão de uma doença em pessoas com sinais clínicos ou sintomas (eletrocardiograma, raios X para detectar fraturas ósseas);
 - Tratamento: visa melhorar ou manter o estado de saúde, evitar uma deterioração maior ou atuar como paliativo;
 - Reabilitação: visa restaurar, manter ou melhorar a função de uma pessoa com uma incapacidade física ou mental.
- Quanto ao estágio de difusão:
 - Futura: em estágio de concepção ou nos estágios iniciais de desenvolvimento;
 - Experimental: quando está submetida a testes em laboratório usando animais ou outros modelos;
 - Investigacional: quando está submetida a avaliações clínicas iniciais (em humanos);
 - Estabelecida: considerada pelos provedores como um enfoque padrão para uma condição particular e difundida para uso geral;
 - Obsoleta/abandonada/desatualizada: sobrepujada por outras tecnologias ou foi demonstrado que elas são inefetivas ou prejudiciais.

Embora não haja uma classificação padrão para tecnologias de saúde, nota-se que o termo de tecnologia de saúde é bem amplo. A próxima sub-seção é apresentada a diferença entre dispositivo médico e aplicativo de saúde.

3.4.2 Diferença entre dispositivo médico e aplicativo de saúde

Segundo a Resolução da Diretoria Colegiada (RDC) 751/2022 da Anvisa [MINISTÉRIO DA SAÚDE, 2022b] que dispõe sobre classificação de riscos, os regimes de notificação e de registro, e os requisitos de rotulagem e instruções de uso de dispositivos médicos, um Dispositivo Médico é qualquer instrumento, aparelho, equipamento, implante, dispositivo médico para diagnóstico *in vitro*, software, material ou outro artigo, destinado pelo fabricante a ser usado, isolado ou conjuntamente, em seres humanos, para algum dos seguintes propósitos médicos específicos, e cuja principal ação pretendida não seja alcançada por meios farmacológicos, imunológicos ou metabólicos no corpo humano, mas que podem ser auxiliados na sua ação pretendida por tais meios:

1. diagnóstico, prevenção, monitoramento, tratamento (ou alívio) de uma doença;
2. diagnóstico, monitoramento, tratamento ou reparação de uma lesão ou deficiência;
3. investigação, substituição, alteração da anatomia ou de um processo ou estado fisiológico ou patológico;
4. suporte ou manutenção da vida;
5. controle ou apoio à concepção; ou
6. fornecimento de informações por meio de exame *in vitro* de amostras provenientes do corpo humano, incluindo doações de órgãos e tecidos.

Já para a OMS, por meio do GHTF [SAÚDE 4.0, 2015], um dispositivo médico é definido como qualquer instrumento, aparelho, equipamento, software, material ou outro produto que possa ser utilizado, sozinho ou combinados, em seres humanos, para um ou mais dos seguintes propósitos médicos específicos:

1. diagnóstico, prevenção, monitoramento, tratamento ou alívio de doenças,
2. diagnóstico, monitoramento, tratamento, alívio ou compensação de uma lesão,
3. investigação, substituição, modificação ou suporte da anatomia ou de um processo fisiológico,
4. apoiando e sustentando a vida,
5. controle da concepção,
6. desinfecção de dispositivos médicos,
7. fornecimento de informações através de exames *in vitro* de amostras derivadas do corpo humano.

De acordo com a OMS, saúde móvel [MAASS et al., 2022] é um termo geral que abrange a saúde pública e a prática médica através do uso de smartphones, sensores, assistentes pessoais digitais, dispositivos de monitoramento sem fio ou outros dispositivos sem fio. Dentre as tecnologias de saúde móvel, os aplicativos móveis de saúde são softwares para smartphones e tablets que podem promover a saúde das pessoas. Esses aplicativos podem ser utilizados para suporte a pessoas com doenças crônicas, melhorar a adesão a tratamentos, aumentar a autonomia dos pacientes, além de oferecerem oportunidades de diminuição de custos sem comprometer a qualidade dos serviços de saúde.

Pelas definições tanto da OMS quanto da Anvisa, um aplicativo móvel de saúde pode ser considerado um dispositivo médico. Segundo a RDC 657/2022 da Anvisa [MINISTÉRIO DA SAÚDE, 2022a] que dispõe sobre a regularização de softwares como dispositivos médicos, um software é considerado um dispositivo médico (Software as a Medical Device - SaMD) caso ele atenda à definição de dispositivo médico, podendo ser de diagnóstico *in vitro* (IVD) ou não, sendo destinado a uma ou mais indicações médicas, e que realizam essas finalidades sem fazer parte de hardware de dispositivo médico. Inclui os aplicativos móveis e softwares com finalidades *in vitro*, se suas indicações estiverem incluídas na definição geral de dispositivos médicos. Incluem-se nesta definição, entre outros, o software licenciado por assinatura e hospedado centralmente (Software as a Service), que se enquadre na definição de dispositivos médicos.

A RDC 751/2022 da Anvisa [MINISTÉRIO DA SAÚDE, 2022b] acrescenta características de um SaMD: o SaMD pode ser executado em uma plataforma computacional de propósito geral (finalidade não médica); a "plataforma computacional" inclui recursos de hardware e software (sistema operacional, hardware de processamento, armazenamento, banco de dados, dispositivos de visualização, dispositivos de entrada, linguagem de programação, etc.); "sem fazer parte de" significa que o programa não precisa do hardware de um dispositivo médico para atingir sua finalidade de uso; um software não é considerado SaMD se seu objetivo é controlar o hardware de um dispositivo médico; um SaMD pode ser usado em combinação (por exemplo, como módulo) com outros produtos, incluindo outros dispositivos médicos; um SaMD pode interagir com outros dispositivos médicos, incluindo hardware de outros dispositivos médicos e outro SaMD, bem como software de uso geral; e os aplicativos móveis de saúde que atendem à definição são considerados SaMD.

Pela definição apresentada de SaMD, nem todos os aplicativos móveis de saúde podem ser considerados um SaMD. A RDC 657/2022 da Anvisa [MINISTÉRIO DA SAÚDE, 2022a], segundo o parágrafo 2 do seu artigo 1, não se aplica aos seguintes softwares: softwares de bem estar (softwares destinados a encorajar e manter o bem-estar, incluindo atividades saudáveis como exercícios físicos, ou a encorajar e manter o controle da saúde e um estilo de vida saudável que não são destinados à prevenção, diagnóstico, tratamento, reabilitação ou anticoncepção); relacionados em lista disponibilizada pela Anvisa de produtos não regulados; utilizado exclusivamente para gerenciamento administrativo e financeiro em serviço de saúde; que processa dados médicos demográficos e epidemiológicos, sem qualquer finalidade clínica diagnóstica ou terapêutica; e embarcado em dispositivo médico sob regime de vigilância sanitária.

A discussão sobre a distinção de aplicativos de saúde categorizados como SaMD daqueles que não são categorizados como SaMD é foco do trabalho apresentado em MAASS [2022], onde os aplicativos médicos correspondem ao SaMD dentro de um e os aplicativos de saúde correspondem aos não SaMD. Nele, são apresentadas as definições de aplicativos médicos e de aplicativos de saúde em diversos países, e evidencia que os SaMD são um grupo particular dos aplicativos de saúde, os quais requerem regulação legal para serem utilizados como dispositivos médicos.

Essa distinção de SaMD como um subgrupo de aplicativos de saúde é importante para a sua adequada regulação. A RDC 848/2024 da Anvisa que dispõe sobre os requisitos essenciais de segurança e desempenho aplicáveis aos dispositivos médicos e dispositivos médicos para diagnóstico *in vitro*, descreve em sua subseção VIII um conjunto de itens a

serem atendidos pelos desenvolvedores de SaMD. Uma vez que aplicativos de saúde podem ser categorizados como SaMD, é importante instrumentos para averiguar se tais aplicativos atendem às resoluções e recomendações estabelecidas. A próxima sub-seção apresenta brevemente os principais aspectos sobre Avaliação de Tecnologias de Saúde, enfatizando a Avaliação de Tecnologias Digitais em Saúde.

3.4.3 Importância e Impacto da Avaliação de Tecnologia em Saúde Digital

Para concretizar os benefícios e minimizar os riscos, é essencial um processo de avaliação robusto para informar os decisores sobre o financiamento e o planejamento de todas as novas tecnologias de saúde, incluindo as intervenções TDS. A avaliação das tecnologias de saúde (ATS) oferece um quadro para avaliar de forma abrangente as novas tecnologias de saúde. Muitos países utilizam habitualmente a ATS para determinar a relação custo-benefício e o impacto orçamental das novas tecnologias de saúde para informar as decisões de financiamento/investimento público.

Assim é preciso a adoção de mecanismos de regulação e de tecnologias que sejam embasadas por evidências científicas. A Lei 12.401 de 2011 [PRESIDÊNCIA DA REPÚBLICA - CASA CIVIL, 2011] estabelece que o uso de evidências de eficácia, segurança e custo-efetividade deve nortear a incorporação de tecnologias no SUS. A Avaliação de Tecnologias em Saúde (ATS) busca subsidiar as decisões quanto à adoção de tecnologias pelos sistemas de saúde. Em Goodman [1998], a ATS é descrita como um campo multidisciplinar de análise de políticas, que estuda as implicações clínicas, sociais, éticas e econômicas do desenvolvimento, difusão e uso da tecnologia em saúde.

Uma vez que tecnologia em saúde abrange uma grande variedade de componentes como medicamentos, procedimentos, instrumentos e equipamentos, é preciso compreender as dimensões a serem avaliadas [ELIAS, 2013]:

- **Eficácia:** benefício das tecnologias em condições ideais de utilização, e os resultados são avaliados pela interpretação de risco relativo, risco absoluto, redução de risco absoluto, e quantidade de vezes que a tecnologia deve ser aplicada para tratamento;
- **Segurança:** existência de eventos adversos provenientes da tecnologia;
- **Efetividade:** benefício das tecnologias em condições reais de utilização dos serviços;
- **Custo-benefício:** análise comparada entre os custos e os resultados em saúde gerados pela tecnologia. Busca estimar em que medida o resultado gerado compensa os custos decorrentes de sua utilização;
- **Impacto no orçamento:** estimativa do aumento ou redução de gastos ao se introduzir e difundir a tecnologia no serviço de saúde;
- **Equidade:** análise da proporção da população a ser beneficiada em razão de suas necessidades sócio-culturais, biológicas e de gênero;
- **Ética:** adequação ao código de princípios de moralidade definidos pela sociedade ou cultura, considerados ideais no caráter e na cultura;
- **Bases Legais:** compatibilidade e adequação à legislação vigente, necessidade de alteração de normas;
- **Logística:** adequação e necessidades de alteração de infraestrutura, de pessoal, de transporte, armazenamento, e todos os aspectos relacionados a cobertura e acesso de qualidade;

- **Macroeconomia:** análise de alocação de recursos no sistema de saúde e dos efeitos nas políticas de propriedade intelectual, na regulação, no investimento em inovação, na transferência de tecnologias e no aumento ou diminuição de empregos;
- **Meio ambiente:** geração de resíduos poluentes, condições e recursos necessários para mitigação de possíveis danos ao meio ambiente.

Essas dimensões normalmente são ponderadas sob a perspectiva de custo-oportunidade, investigando as consequências de curto e médio prazos, avaliando os efeitos desejados, indesejados e incertezas em torno da adoção da tecnologia.

As tecnologias digitais de saúde (TDS) que permitem a monitorização remota ou a autogestão dos pacientes oferecem muitos benefícios potenciais, especialmente para os pacientes que sofrem de doenças crônicas. Exemplos de potenciais benefícios incluem a redução dos tempos de viagem e de espera clínica, redução da exposição e propagação de infecções, ganhos de confiança nas condições de autogestão e melhor ligação do paciente com a sua equipa de saúde em momentos críticos.

No entanto, a utilização de TDSs como intervenção de saúde pode representar uma mudança significativa na prestação de serviços de saúde. Os benefícios só podem ser obtidos se o pessoal for apoiado por facilitadores apropriados, por exemplo, recursos adequados e apoio à gestão de mudanças, interoperabilidade de sistemas e gestão adequada da qualidade dos dados, financiamento apropriado e defensores da nova intervenção dentro da organização

Além disso, com os benefícios potenciais vêm riscos exclusivos dos TDSs, como ameaças à privacidade do paciente devido à segurança cibernética inadequada, uso de informações incorretas do paciente devido à má interoperabilidade/gestão da qualidade dos dados, falta de acesso em áreas de conectividade/infraestrutura deficiente e ameaças éticas. como uma falsa sensação de segurança por ser monitorado ou interpretação incorreta dos resultados dos testes sem supervisão médica [MAAß, FREYE, PAN... 2022]. Além disso, os TDSs geralmente coletam uma grande quantidade de dados pessoais, e os TDSs que promovem mudanças comportamentais costumam usar esses dados para desenvolver algoritmos de comportamento preditivo. As empresas de tecnologia também podem usar transmissões de dados com identificadores de usuário vinculáveis de TDSs para fins indesejáveis de marketing e análise.

As ameaças à privacidade resultantes da forma como os TDSs são concebidas são desafios únicos para avaliar os riscos versus benefícios destas tecnologias. Estes riscos únicos devem ser considerados juntamente com os riscos comuns a outras tecnologias, por exemplo, tratamentos menos eficazes, diagnósticos errados e classificação errada da gravidade da doença. Na próxima subseção são abordados brevemente os processos e protocolos de ATS.

3.4.4 Processos e Protocolos de Avaliação de Tecnologia em Saúde Digital

As primeiras iniciativas para estabelecer formas de avaliação para Tecnologias Digitais em Saúde (TDS) estavam inicialmente centradas na telemedicina. Foram desenvolvidas propostas de protocolos baseados em sistemas de pontuação para avaliar a qualidade dos estudos, com adaptações das avaliações ao contexto da telemedicina, englobando participantes de diversas disciplinas relevantes. Por meio do financiamento da Comissão Europeia, foi desenvolvido um quadro específico para a telemedicina, chamado Modelo

de Avaliação de Aplicações de Telemedicina (MAST). Esse modelo incorporou o EUnetHTA (European Network for Health Technology Assessment) Model Core, que é uma rede estabelecida para criar uma estrutura eficaz e sustentável para avaliações de tecnologias em saúde em toda a Europa [EUNETHTA, 2016]. O EUnetHTA Model Core possui nove domínios, com 51 temas e 145 questões. O conteúdo das estruturas incluídas foi mapeado para as 145 questões do Modelo Principal de ATS num processo de duas fases. Inicialmente, foram incluídos tópicos e questões específicas do TDS levantadas pelas estruturas, mas ainda não incluídas no modelo, para garantir uma compilação abrangente do conteúdo do TDS. Para novos tópicos específicos do TDS, novos nomes de tópicos foram propostos, e para novos problemas específicos do TDS, novos identificadores de problemas foram atribuídos usando um prefixo TDS. Posteriormente, todo o conteúdo recomendado por cada framework foi mapeado para o conjunto ampliado de questões.

Por sua vez, a Organização Mundial da Saúde (OMS) produziu o Kit de Ferramentas da Estratégia Nacional de eSaúde em 2012 [OMS, 2012], uma vez que vários países solicitaram orientações acerca do uso de tecnologias de eSaúde. Essas iniciativas resultaram em 2019 na Diretriz da OMS sobre Intervenções Digitais para o Fortalecimento do Sistema de Saúde. No entanto, tal diretriz não contempla as TDSs destinadas ao uso individual do paciente.

No âmbito da regulamentação, o Fórum Internacional de Reguladores de Dispositivos Médicos (IMDRF) estabeleceu um grupo de trabalho sobre "Software como Dispositivo Médico" (SaMD), publicando orientações sobre definições, categorização de riscos, gestão de qualidade e avaliação clínica [IMDRF, 2014]. Estas diretrizes foram adotadas pela Coordenação de Dispositivos Médicos da União Europeia em 2019, ampliando o escopo da regulamentação para todas as tecnologias digitais com finalidade médica.

Apesar desses esforços, havia poucas orientações sobre os critérios de avaliação para o financiamento público das TDSs, embora algumas agências europeias de Avaliação de Tecnologias em Saúde tenham publicado critérios de lista de verificação para avaliar aplicações de saúde em seus sites. O quadro Digi-HTA foi implementado na Finlândia para avaliações de TDSs, enquanto a Portaria sobre Aplicações de Saúde Digital agora regula o reembolso de aplicações de saúde digitais na Alemanha. No entanto, nenhum desses quadros recentes abrange todos os domínios recomendados pelo modelo principal da EUnetHTA e não são informados por estudos sobre as preferências das partes interessadas ao tomar decisões de financiamento/reembolso público sobre TDSs.

O trabalho apresentado em [VON HUBEN et al., 2021], apresenta uma lista de propostas de protocolos de avaliação de tecnologias digitais em saúde. Dada a complexidade do tema, reconhecida pela OMS e Comissão Europeia, o restante desta subseção dedica-se a instrumentos de avaliação de aplicativos de saúde. Especificamente para aplicativos móveis de saúde, existem protocolos específicos de avaliação que serão descritos a seguir.

A *Mobile App Rating Scale* (MARS) é um protocolo padronizado para avaliar a qualidade de aplicativos móveis, especialmente na área da saúde. A MARS contém 23 itens divididos em cinco subescalas que analisam diferentes aspectos dos aplicativos: Engajamento que avalia o quão interessante e atraente o aplicativo é para os usuários, considerando envolvimento, entretenimento, personalização, interatividade e adequação

ao público-alvo; Funcionalidade que avalia o desempenho do aplicativo, navegação, design gestual e facilidade de uso; Estética que mede o apelo visual do aplicativo, incluindo design e gráficos; Qualidade da informação que avalia a precisão e integridade das informações fornecidas; Qualidade subjetiva: Avalia a satisfação geral do usuário, incluindo a intenção de usá-lo novamente e recomendá-lo a outros [STOYANOV et al., 2015]. A MARS é amplamente utilizada para avaliar aplicativos em várias áreas da medicina, como cardiologia, reumatologia e obstetria. Também é empregada no design e desenvolvimento de novos aplicativos de saúde. Além disso, a MARS avalia a segurança dos aplicativos em termos de precisão das informações e privacidade do usuário.

A *Adapted MARS (A-MARS)* é uma versão adaptada da MARS, específica para aplicativos e ferramentas eletrônicas relacionadas à saúde. Inclui novas seções para avaliar: Se o aplicativo ou ferramenta eletrônica oferece recursos atuais e relevantes; Estratégias recomendadas relacionadas à área de preocupação; Soluções para sintomas identificados; Capacidade de uso em tempo real; Facilidade de acesso à ajuda ou suporte [ROBERTS et al., 2021]. O *User Version of MARS (uMARS)* é uma versão simplificada da MARS, focada no usuário final. Contém 20 itens divididos em seis subescalas: engajamento, funcionalidade, estética, qualidade da informação, qualidade subjetiva e impacto percebido pelos usuários. O uMARS é útil para obter a percepção dos usuários sobre a qualidade dos aplicativos durante o desenvolvimento e teste, contribuindo para melhorias gerais [STOYANOV et al., 2016].

A Escala de Usabilidade do Sistema (SUS) é um protocolo de avaliação de usabilidade para uma ampla gama de produtos, incluindo aplicativos móveis. Desenvolvida em 1986, a SUS avalia a efetividade, eficiência e satisfação da experiência do usuário, mas não identifica problemas específicos de usabilidade [ZHOU et al., 2019]. O PSSUQ, desenvolvido pela IBM, avalia as percepções dos usuários sobre a usabilidade de um sistema após o uso. Com 19 itens, mede utilidade do sistema, qualidade da informação e da interface de usuário [FRUHLING; LEE, 2005]. O *mHealth App Usability Questionnaire (MAUQ)* foi projetado para avaliar a usabilidade de aplicativos móveis de saúde. Com 30 questões, ele avalia: a Facilidade de através de questões sobre navegação, legibilidade e clareza das instruções; Utilidade por meio de questões sobre benefícios percebidos e utilidade para gerenciar e rastrear informações de saúde; Satisfação com questões sobre design, funcionalidade e experiência geral; Apelo visual com questões sobre esquema de cores e design geral; Qualidade da informação com questões sobre precisão, relevância e abrangência das informações [ZHOU et al., 2019].

O *Mobile Health Evidence Reporting and Assessment (mERA)* é uma escala com 16 itens focada em relatórios sobre intervenções de saúde móvel. Promove a clareza e completude na divulgação de pesquisas envolvendo ferramentas móveis na área da saúde, abordando especificações técnicas necessárias para um relatório completo [AGARWAL et al., 2016]. A ISO/TS 82304-2 é uma especificação técnica da ISO que orienta o desenvolvimento e avaliação de software de saúde, com foco na usabilidade. Avalia eficácia, eficiência e satisfação do usuário, distribuindo a avaliação entre 82 itens que cobrem aspectos como acessibilidade, segurança dos dados e robustez técnica [ISO 82304-2, 2021]. O ARIA (*Alberta Rating Index for App*) é um protocolo composto por 06 itens para avaliar a apresentação do aplicativo pelo hospedeiro para ajudar o usuário a decidir pela instalação ou não do aplicativo, e 12 itens para avaliar a qualidade do conteúdo e usabilidade do aplicativo após a instalação. Foram desenvolvidas duas versões

do ARIA, uma voltada para os usuários de aplicativos e outra para o provedor de serviços de saúde [AZAD-KHANEHGAH; RODUTA ROBERTS; LIU, 2022].

Essas ferramentas ajudam desenvolvedores, pesquisadores e profissionais de saúde a avaliar e melhorar a qualidade, segurança e usabilidade de aplicativos móveis de saúde.

3.4.5 Análise do Resultado da Avaliação de Tecnologia em Saúde Digital

A análise dos resultados de avaliação de uma tecnologia de saúde visa garantir a confiabilidade e a própria validade a que se propõe a medir, monitorar e gerar informações sobre variáveis ou condições do usuário.

A avaliação prévia por especialistas das áreas envolvidas em tecnologia e saúde é substancial. Assim, a busca por especialistas pode ser realizada através de consultas em bases de currículos, plataformas de redes de profissionais, entre outras, cabendo a ressalva da importância da descrição dos critérios de seleção dos especialistas bem como da descrição das qualificações dos mesmos na seção de resultados do protocolo [ALEXANDRE; COLUCI, 2011]. A norma ABNT/ISO/IEC 25062-2011 [2011], recomenda uma amostragem mínima de oito avaliadores para cada categoria de usuários, a fim de garantir representatividade em relação ao grupo de usuários pretendidos. Para a obtenção dos dados, neste rol de avaliadores, sugere-se como método, a Técnica Delphi, que consiste em obter consenso entre os especialistas – chamados de “juízes – através de rodadas entre as quais realiza-se adequações nas questões a partir dos feedbacks dos “juízes” [KEENEY; HASSON; MCKENNA, 2011]. Na Técnica Delphi, o número de rodadas é determinado visando um consenso final, através dos ajustes das divergências justificados pelas sugestões dos “juízes”. Assim, na primeira rodada, os especialistas respondem quantitativamente a cada questão, acompanhada de justificativas qualitativas. Após a obtenção dos resultados, os especialistas participam de novas rodadas, sucessivamente, repetindo o processo até que o consenso seja alcançado sobre o objeto de estudo. Para que o processo não se estenda por muitas rodadas, têm-se optado pela realização de uma versão modificada da Técnica Delphi, limitando o número de rodadas de duas a quatro, com o intuito de gerenciar o tempo disponível para a construção do consenso [ZARILI et al., 2021].

Para a avaliação quantitativa das apreciações dos “juízes” em busca do consenso final, um método indicado é o Índice de Validade de Conteúdo (IVC), muito utilizado na área de saúde na avaliação de constructos [ALEXANDRE; COLUCI, 2011]. O IVC mede a proporção ou porcentagem de concordância entre os especialistas, permitindo uma análise individual de cada item do instrumento e posterior análise global.

O método consiste em indicar para o especialista qual característica deverá ser analisada para cada item ou constructo e classifica-la numa escala de Likert com pontuação de 1 a 4, tomando como exemplo a acessibilidade, através da Questão X: “O aplicativo possui os recursos de acessibilidade”:

- 1 – NÃO POSSUI;
- 2 – POSSUI, MAS INADEQUADO;
- 3 – POSSUI, ADEQUADO;
- 4 – POSSUI, TOTALMENTE ADEQUADO.

O cálculo do índice é realizado pela soma do número dos itens que receberam “3” ou “4” dividida pelo número total de respostas [POLIT; BECK, 2006] [ALEXANDRE; COLUCI, 2011]. Juntamente com as justificativas dos especialistas, os itens que receberam notas “1” ou “2” devem ser ajustados ou retirados do instrumento para nova rodada [ALEXANDRE; COLUCI, 2011]. Os cálculos demonstrados até aqui são para a análise individual de cada item do instrumento. Para a análise global, embora não haja consenso na literatura, uma opção prática é calcular o IVC médio entre os itens que receberam avaliações “3” e “4” [YUSOFF, 2019].

A literatura indica que o IVC individual aceitável é de 0,78 e, no mínimo, 0,80 para o IVC global, porém, valores acima de 0,90 são consideravelmente excelentes para um instrumento [YUSOFF, 2019].

Uma outra opção para a análise da concordância entre especialistas é o coeficiente de Kappa de Cohen, que não somente avalia concordância entre os especialistas, mas também a confiabilidade e reprodutibilidade entre os observadores (no caso, os especialistas). O Kappa de Cohen permite avaliar as respostas de apenas dois observadores, o que na versão modificada, o Kappa de Fleiss, pode ser estendida para um número maior de avaliadores [MCHUGH, 2012]. O coeficiente do teste de Kappa (Cohen ou Fleiss) varia de 0 a 1, sendo valores próximos ao zero indicam nenhuma concordância ou confiabilidade o que no caso contrário indica alta confiabilidade [LANDIS; KOCH, 1977].

A confiabilidade e a concordância são quesitos importantes na classificação, no desenvolvimento de escalas e instrumentos visando garantir a qualidade e validação dos protocolos, fornecendo informações sobre o erro inerente a qualquer diagnóstico, pontuação ou medição, na qual a quantidade de erro de medição pode determinar a validade dos resultados ou pontuações obtidas [KOTTNER et al., 2011].

Para verificar a consistência interna dos resultados, ou seja, o quanto as questões estão correlacionadas para atender e medir determinado constructo, sugere-se o coeficiente Alfa de Cronbach. Seu cálculo considera a variância dos valores de cada questão respondida e a variância de cada indivíduo respondente. Naturalmente, o coeficiente Alfa de Cronbach varia entre 0 e 1, sendo o valor mínimo aceitável para o alfa é 0,70.

A consistência interna dos itens da escala ou instrumento abaixo desse valor, indica uma confiabilidade interna fraca, necessitando de ajustes consideráveis. Quanto mais os valores se aproximam de 1, maior a consistência interna. Contudo, valores muito próximos de 1 podem indicar a presença de redundância, ou seja, vários itens podem estar medindo o mesmo elemento [STREINER, 2003]. Uma vez passado pelas análises anteriores, pode-se ampliar para uma amostragem de usuários da tecnologia e, os dados coletados podem, então, ser analisados para fornecer insights sobre a usabilidade, satisfação e eficácia do aplicativo.

3.5 Desenvolvimento, validação e aplicação do protocolo

A fim de garantir a qualidade de um produto de software, foi adotado um protocolo de desenvolvimento fundamentado nos padrões estabelecidos pela ISO 25000, que integra as diretrizes previstas tanto na ISO/IEC 9126 quanto na ISO/IEC 14598 [ISO 14598, 1998] [ISO 9126, 2001] [ISO 25000, 2014].

3.5.1 Processo de desenvolvimento do Protocolo de Avaliação

A ISO 25000, cria um conjunto de regras, que envolve dois processos principais: especificação de requisitos e a avaliação da qualidade. A especificação de requisitos, envolve desde a análise comparativa de métodos existente na literatura, quanto o objetivo, configuração, medidas e níveis de pontuação e critérios de julgamento. A avaliação da qualidade abrange a validação da estrutura do protocolo quanto a eficácia do protocolo.

3.5.2 Especificação de requisitos

Para a especificação dos requisitos, é composto por cinco etapas:

- (i) análise e comparação de métodos de avaliação de aplicativos de saúde: identificação das características;
- (ii) definição dos objetos e contextualização da avaliação do protocolo;
- (iii) definição e configuração do modelo de qualidade do protocolo;
- (iv) identificação das medidas da avaliação do protocolo;
- (v) definição de níveis de pontuação e critérios de julgamento do protocolo.

3.5.2.1 Análise e comparação de métodos de avaliação de aplicativos de saúde

Para a etapa análise e comparação de métodos de avaliação de aplicativos de saúde, é feita uma revisão sistemática da literatura para a identificação das características de cada um dos métodos de avaliação.

Com o intuito de compreender e verificar quais as características estão presentes em cada método de avaliação, identificam-se todas as características de cada método e elaboram-se um quadro comparativo. Um quadro comparativo é uma ferramenta valiosa para a tomada de decisões, permitindo a análise de semelhanças e diferenças entre entidades [GIACOMETTI; MARKHOFF; SOULET, 2021]. Neste caso específico, é possível identificar as características presentes, assim como as ausentes, dentro de cada método pesquisado.

Para um método de avaliação atualizado é necessário abranger características gerais e específicas, conforme o objetivo. Além destes, a qualidade subjetiva também é levada em consideração por referir-se à percepção individual de qualidade, muitas vezes relacionada a aspectos de experiências individuais.

3.5.2.2 Definição dos objetos e contextualização da avaliação do protocolo

Na etapa de definição dos objetos e contextualização da avaliação do protocolo são inicialmente divididas em quatro grandes áreas. Cada uma dessas áreas tem subdivisões em categorias, as quais, posteriormente, constituem as categorias a serem avaliadas dentro do protocolo. É essencial também definir os objetivos específicos de extração dos dados de cada categoria antes da elaboração do protocolo.

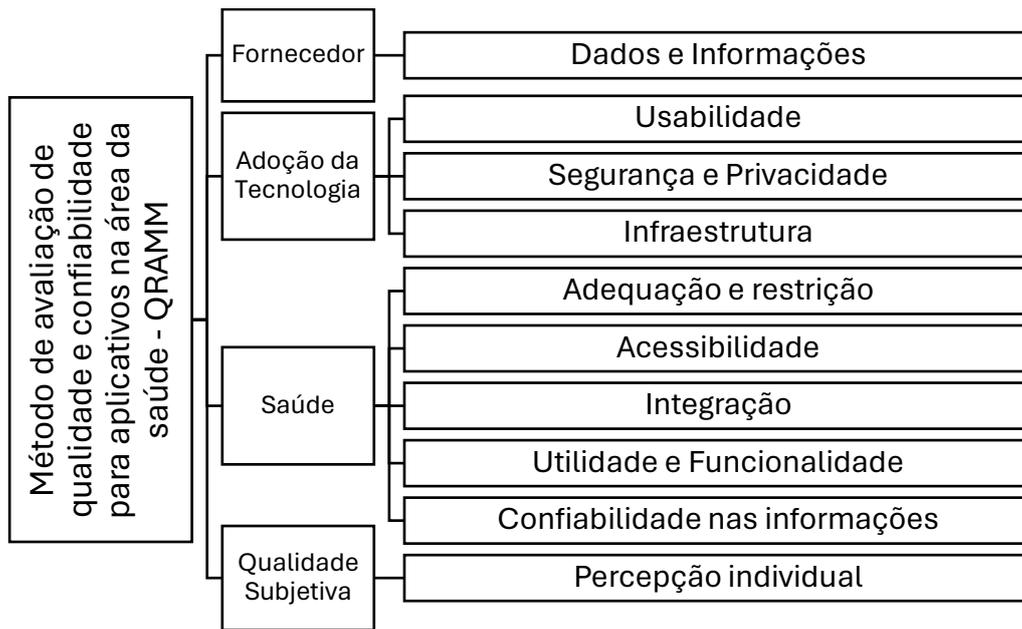


Figura 3.9 - Visão esquemática do Método de Avaliação para Aplicativos na área da Saúde

Considerando a importância de manter a continuidade do pensamento durante a avaliação de um aplicativo, é fundamental estabelecer uma ordem adequada para as categorias. A sequência das categorias deve ser organizada de forma a promover uma fluidez no pensamento e na compreensão das questões, a fim de obter respostas mais coerentes e precisas. Dessa forma, a ordenação das categorias deve ser cuidadosamente planejada, levando em consideração a interdependência dos aspectos avaliados e a progressão natural das áreas de avaliação [FERAUCHE; SILVA; ITO, 2024].

A ordem das categorias em um formulário de avaliação de aplicativos de saúde pode variar dependendo da abordagem e prioridades específicas de cada criador. No entanto, a sequência melhor analisada é a seguinte:

- **Informações iniciais:** Avaliar as informações sobre funcionalidade, uso, fornecedor do aplicativo, atualizações e fonte de download.
- **Acessibilidade:** Avaliar recursos de acessibilidade, suporte para pessoas com deficiências especiais e conformidade com as diretrizes de acessibilidade, como as WCAG.
- **Adequação e restrição:** Avaliar classificação indicativa, informações sobre riscos e contraindicações, atendimento às necessidades e restrições do usuário, orientações claras e capacidade de uso imediato.
- **Usabilidade:** Avaliar organização das informações, facilidade de recuperação de erros, consistência na navegação, acessibilidade às funções e intuitividade das interações.
- **Segurança e privacidade:** Avaliar medidas de segurança, proteção de dados, transparência no tratamento das informações e conformidade com regulamentações de privacidade.
- **Utilidade e funcionalidade:** Avaliar gerenciamento de saúde, acesso a serviços, desempenho do aplicativo e capacidade de personalização para atender às necessidades individuais dos usuários.

- **Integração com outros sistemas de saúde e/ou profissional de saúde:** Avaliar capacidade de compartilhamento de informações, integração com dispositivos e sistemas existentes e facilidade de comunicação e colaboração entre profissionais e usuários.
- **Confiabilidade das informações:** Avaliar precisão, clareza, base científica, linguagem utilizada e processo de revisão das informações fornecidas pelo aplicativo.
- **Infraestrutura:** Avaliar segurança das informações dos pacientes, sincronização, integração com sistemas de comunicação, proteção de dados e compatibilidade com dispositivos.
- **Qualidade Subjetiva:** Avaliar o grau de satisfação subjetiva do usuário com o aplicativo.

Seguir uma sequência lógica e coerente permite que os avaliadores conduzam a avaliação de maneira mais eficiente e facilita a compreensão das informações coletadas [FERAUCHE; SILVA; ITO, 2024].

3.5.2.3 Definição e configuração do modelo de qualidade do protocolo

Para a definição e configuração do modelo de qualidade do protocolo, busca atribuir um valor às características de qualidade e confiabilidade estabelecidas pelo protocolo. É essencial realizar uma medição adequada, portanto essa abordagem permite quantificar e interpretar a qualidade do critério a ser avaliado, fornecendo uma compreensão mais precisa do desempenho do aplicativo de saúde. As medidas são definidas mapeando as categorias em atributos mensuráveis, permitindo assim que elas sejam medidas e pontuadas.

Com o propósito de atingir esse objetivo, é escolhida comumente a abordagem da utilização de questionários como técnica de coleta de dados. Os questionários de coleta de dados são um método comum usado em pesquisas para coletar informações sobre variáveis-alvo. Eles podem ser administrados por vários meios, como métodos baseados em papel ou coleta eletrônica de dados (EDC) [MARSHALL, 2005].

As abordagens de EDC, que envolvem o uso de dispositivos eletrônicos para coletar e armazenar dados, oferecem vantagens sobre os métodos baseados em papel, incluindo pontuação e interpretação automatizadas, economia de tempo e custos e maior eficiência [MHIMDI; KARAA; GHEZALA, 2022]. Essa técnica envolve a elaboração de um conjunto de perguntas objetivas, que possibilitam a coleta de informações por meio de respostas claras e concisas.

Ainda nessa etapa, é estudado e definido a forma de coleta dos dados de cada uma das questões.

As escalas de medição em pesquisas são importantes para determinar a natureza das relações entre variáveis e orientar análises estatísticas apropriadas [DALATI, 2018] [STEVENS, 1946]. A classificação de escalas de Stevens [1946] enfatiza a importância de compreender o nível de medição, como nominal, ordinal, intervalo e razão, na seleção de métodos estatísticos. A distinção entre escalas intervalares e ordinais depende da crença na linearidade das relações entre medidas e quantidades reais de atributos, influenciando a escolha entre análises paramétricas e não paramétricas. Compreender as escalas de medição é fundamental para a interpretação precisa dos dados e a tomada de decisão estatística apropriada na pesquisa. Os tipos de escalas de medição incluem escalas

Likert, Guttman, diferencial semântico, classificação, Thurstone e Borgadus [AINI; ZULIANA; SANTOSO, 2018].

- **Escala Likert:** desenvolvida pelo psicólogo Rensis Likert, na década de 1930, e é amplamente utilizada em pesquisas e questionários. A escala Likert geralmente consiste em uma série de afirmações ou itens aos quais os respondentes devem indicar seu grau de concordância ou discordância. Essa escala é composta por uma série de níveis de resposta que variam de acordo com o número de opções oferecidas. A escala pode variar de 3 pontos (por exemplo, concordo, neutro, discordo) até 7 pontos (por exemplo, concordo plenamente, concordo moderadamente, neutro, discordo moderadamente, discordo plenamente) [BARUA, 2013] [PERES; BUENO; PIASSON, 2016].
- **Escala Guttman:** desenvolvida pelo sociólogo Louis Guttman, que era conhecido por suas contribuições significativas para a teoria da medição e a análise de dados na ciência social. Ele propôs essa técnica de medição na década de 1940 como uma maneira de medir e classificar a ordem de itens ou conceitos de acordo com sua dificuldade ou severidade. Muito parecida com a escala de Likert, porém a principal diferença entre a escala de Guttman e a de Likert é que na escala de Guttman as afirmações selecionadas incorporam a ideia da afirmação anterior, ou seja, é uma escala acumulativa. O respondente só responde questões posteriores se os escores das questões anteriores satisfizerem um determinado nível pré-estabelecido ou considerado limite. A Escala Guttman é frequentemente usada em pesquisas sociais, psicológicas e educacionais para avaliar atitudes, opiniões e comportamentos dos respondentes [PERES; BUENO; PIASSON, 2016].
- **Diferencial Semântico:** é uma técnica utilizada em pesquisa para medir a opinião ou atitude dos respondentes em relação a um determinado objeto, conceito ou ideia. Desenvolvida por Charles Osgood na década de 1950, essa escala busca capturar a valência, intensidade e direção das emoções ou avaliações associadas ao objeto em questão. A estrutura básica da escala consiste em uma série de pares de adjetivos antônimos, geralmente colocados em uma escala linear, em que os respondentes são solicitados a marcar um ponto que melhor represente sua opinião ou sentimentos. Por exemplo, os extremos da escala podem incluir adjetivos como "bom" e "ruim", "positivo" e "negativo", "satisfatório" e "insatisfatório". Os respondentes são instruídos a selecionar um ponto na escala que melhor descreva sua percepção em relação ao objeto em análise. A distância entre os pontos escolhidos pode refletir a intensidade da atitude ou sentimento do respondente. Além disso, a direção em que o ponto é marcado indica se a avaliação é favorável ou desfavorável. Essa escala é amplamente utilizada em pesquisas de marketing, psicologia, ciências sociais e outras áreas para avaliar atitudes em relação a produtos, marcas, serviços, personalidades e conceitos abstratos. A precisão e validade dos resultados dependem da seleção cuidadosa dos adjetivos antônimos, além da clareza das instruções fornecidas aos respondentes [PASSOS, 2015].
- **Classificação:** é uma técnica comum de mensuração usada em pesquisas e avaliações. Nesse método, os respondentes são solicitados a classificar os itens de acordo com um critério específico, geralmente em uma escala

ordinal. Essa escala permite uma rápida coleta de dados e é relativamente simples de administrar e entender. No entanto, ela pode não fornecer informações detalhadas sobre a intensidade ou a direção das opiniões dos participantes, já que geralmente é limitada a uma série de categorias pré-definidas. Além disso, a interpretação dos resultados pode ser influenciada pela forma como as categorias são definidas e pela ordem em que são apresentadas aos participantes. A diferença principal entre a escala de classificação e a escala de Likert está na forma como são estruturadas e como os respondentes são solicitados a responder. Na escala de classificação, os respondentes são solicitados a classificar os itens de acordo com um critério específico em uma escala ordinal. Na escala de Likert, os respondentes são apresentados a uma série de afirmações ou itens e são solicitados a indicar seu nível de concordâncias com cada item em uma escala de múltiplos pontos: como “discordo totalmente”, “discordo”, “neutro”, “concordo” e “concordo totalmente”, com diferentes pontos atribuídos a cada opção.

- **Thurstone:** desenvolvida por Louis Leon Thurstone na década de 1920, é uma técnica de medição de atitudes que visa quantificar o grau de concordância ou discordância dos participantes em relação a uma série de afirmações ou itens. Esta escala é construída com base no pressuposto de que as pessoas podem diferenciar entre diferentes graus de atitude ou opinião. A principal característica é a utilização de uma série de afirmações que representam uma gama de atitudes, que são classificadas pelos participantes em termos de concordância ou discordância. Os itens são geralmente selecionados e organizados em uma ordem gradual de acordo com o nível de aceitação social. Após a seleção dos itens, um grupo de juízes especialistas atribui uma pontuação a cada um deles com base em sua objetividade e representatividade em relação ao construto que está sendo medido. Em seguida, os participantes são apresentados a uma série dessas afirmações e solicitados a classificá-las de acordo com sua concordância ou discordância. Uma vez que todas as respostas são coletadas, as pontuações atribuídas pelos juízes são usadas para calcular uma média ponderada para cada item, resultando em uma pontuação média de concordância ou discordância para cada afirmação. Essas pontuações são então usadas para ordenar as afirmações em uma escala unidimensional, representando a intensidade da atitude dos participantes em relação ao construto em questão. A escala de Thurstone é amplamente utilizada em pesquisas de opinião, estudos de mercado e pesquisas sociais para medir atitudes em relação a questões específicas. É uma ferramenta valiosa para capturar nuances nas opiniões dos participantes e fornecer insights detalhados sobre as atitudes de um grupo em relação a determinados tópicos [PASSOS, 2015].
- **Bogardus:** também conhecida como Escala de Borg, é uma ferramenta de avaliação subjetiva da percepção de esforço durante o exercício físico. Ela foi desenvolvida por Gunnar Borg na década de 1980 e é amplamente utilizada em contextos de monitoramento da intensidade do exercício em estudos científicos e na prática clínica. Esta escala é uma técnica simples e eficaz para que os indivíduos avaliem a sensação subjetiva de esforço percebido durante o exercício. Ela consiste em uma série de números ou palavras que

representam uma graduação de esforço, que vai desde "nenhum esforço" até "máximo esforço". Geralmente, os participantes são instruídos a escolher o número ou palavra que melhor descreve a intensidade do esforço percebido no momento. A Escala de Borg é uma ferramenta subjetiva, pois depende da percepção individual do esforço pelo participante, levando em consideração fatores como a sensação de cansaço, dificuldade respiratória, desconforto muscular e ritmo cardíaco percebido. Por isso, ela pode variar de pessoa para pessoa e mesmo para uma mesma pessoa em diferentes situações. Uma das vantagens dessa escala é a sua simplicidade de uso e aplicação, sendo fácil para os participantes compreenderem e utilizarem durante o exercício. Além disso, ela pode ser adaptada para diferentes populações e tipos de exercício, tornando-se uma ferramenta versátil para monitorar a intensidade do esforço percebido em uma ampla gama de situações. No entanto, é importante ressaltar que a Escala de Borg é uma medida subjetiva e, portanto, pode não ser tão precisa quanto medidas objetivas de intensidade do exercício, como a frequência cardíaca ou o consumo de oxigênio. Assim, ela geralmente é utilizada em conjunto com outras medidas para fornecer uma avaliação mais abrangente da intensidade do exercício [SANTOS et al., 2014].

A escala de Borg e a escala de Thurstone são ferramentas de medição subjetiva que dependem da percepção individual dos participantes. Embora sejam úteis em contextos específicos, como avaliação de esforço físico ou opiniões pessoais sobre determinado tema, podem ser eliminadas na criação de um protocolo para a avaliação. Essas escalas podem não ser adequadas para avaliar aspectos específicos dos sistemas, como usabilidade, acessibilidade, eficácia ou confiabilidade. Elas são mais adequadas para avaliações subjetivas de sensações físicas ou opiniões pessoais. As escalas de Borg e Thurstone podem gerar dados difíceis de interpretar devido à natureza subjetiva das respostas. Isso pode dificultar a análise e a tomada de decisões com base nos resultados obtidos. Portanto, ao criar um protocolo para a avaliação de sistemas, é preferível utilizar métodos de avaliação mais objetivos e padronizados [SANTOS et al., 2014].

3.5.2.4 Identificação das medidas da avaliação do protocolo

Nesta etapa estabelecem-se critérios de avaliação, determinando o nível específico e objetivo da avaliação do protocolo. Essa definição possibilitará a comparação entre diferentes produtos, ao mesmo tempo em que identifica oportunidades de aprimoramento.

Para definir o nível específico e objetivo da avaliação do protocolo é necessário atribuir valores quantitativos. Os valores quantitativos podem ser obtidos com a combinação de perguntas e respostas. As pontuações são atribuídas a cada pergunta. Essas perguntas contemplam as categorias definidas na etapa de definição dos objetos e contextualização da avaliação do protocolo (3.6.2.2).

As respostas do questionário podem ser de escolha única ou múltipla, dependendo do objetivo da pergunta. Com a escolha definida é aplicado a escala escolhida em definição e configuração do modelo de qualidade do protocolo (3.6.2.3). Cada questão é elaborada cuidadosamente, incluindo diferentes opções de resposta, a fim de fornecer uma compreensão abrangente das avaliações. A variedade de respostas disponíveis permite capturar de forma mais precisa as opiniões e percepções dos avaliadores.

3.5.2.5 Definição de níveis de pontuação e critérios de julgamento do protocolo

Na última etapa, é definida a fórmula para resolução de cada categoria. A partir dos resultados obtidos para cada categoria, é possível calcular o nível de qualidade para cada uma das categorias definidas na etapa de definição dos objetos e contextualização da avaliação do protocolo.

3.5.3 Colocando em prática o protocolo de avaliação

Como prática de elaboração de protocolo de avaliação, segue como exemplo, o método QRAMHe de FERAUCHE; SILVA; ITO [2024]. Este foi proposto para avaliar aplicativos de saúde quanto a informações iniciais; adequação e restrição; usabilidade; segurança e privacidade; utilidade e funcionalidade; integração com outros sistemas de saúde e/ou profissional de saúde; confiabilidade das informações; infraestrutura; acessibilidade e qualidade subjetiva. Apesar de novo, este método de avaliação encontra-se disponível em <https://qramhe.org/>. Este método de avaliação, conforme FERAUCHE; SILVA; ITO [2024], segue sua estruturação segundo as normas ISO/IEC 25000 [2014], que unifica as normas ISO/IEC 9126 [2001] e ISO/IEC 14598 [1998].

As quatro etapas, segundo a ISO 25000 [2014], serão descritas separadamente a seguir.

3.5.3.1 Definição dos Objetos e Contextualização da Avaliação do Protocolo

Conforme proposto por FERAUCHE; SILVA; ITO [2024], a preparação do método de avaliação seguiu com a nomeação de protocolo QRAMHe; visto que este ainda não está pronto e habilitado para ser uma avaliação.

Inicialmente, para os objetos e contextualização da avaliação do protocolo QRAMHe, foi fundamentada com uma análise da revisão da literatura. O modelo do protocolo QRAMHe foi 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, dentro de cada área [FERAUCHE; SILVA; ITO, 2024]:

- **Área Fornecedor:** foi composta pela categoria Dados e Informações; que foram consideradas as informações iniciais sobre o aplicativo, atualização e descrição do aplicativo e plataformas de distribuição;
- **Área Adoção da tecnologia:** foi dado maior ênfase a facilidade de uso, interface amigável e em ter funcionalidades claras. Além de garantir a segurança e a privacidade dos dados daqueles que utilizam o aplicativo;
- **Área Saúde:** com quatro domínios: adequação e restrição; integração; utilidade e funcionalidade e, por fim, confiabilidade nas informações. A integração tem o propósito das trocas de informações entre o aplicativo e outros relacionamentos, sejam estes outros sistemas ou o profissional de saúde. Com a finalidade de verificar se o aplicativo condiz com o proposto além de verificar se há restrições por meio de compensações monetárias, o domínio de utilidade e funcionalidade estão dispostos para esse tipo de avaliação. Para garantir a autenticidade das informações dispostas no aplicativo, o último domínio desta área, é verificado se há confiabilidade nas informações.

- **Área Qualidade Subjetiva:** refere-se à percepção individual de qualidade que pode variar de pessoa para pessoa. É uma medida subjetiva baseada em opiniões, preferências e experiências pessoais. É importante reconhecer que a qualidade subjetiva não é uma medida objetiva e pode variar de acordo com a perspectiva e o contexto de cada indivíduo.

Com a definição dos objetos e contextualização da avaliação do protocolo avança com a definição e configuração do modelo de qualidade do protocolo.

3.5.3.2 Definição e Configuração do Modelo de Qualidade do Protocolo

Para o protocolo QRAMHe, foi utilizado o questionário, que foram definidas as categorias em atributos mesuráveis, permitindo assim que elas fossem medidas e pontuadas [FERAUCHE; SILVA; ITO, 2024].

Ao final do esboço do protocolo, foi adotado o questionário composto por 50 questões, abordando as 10 categorias, utilizando a Escala Likert com variação de 3 a 5 pontos. A pontuação total pode ser calculada somando-se as respostas de cada item, permitindo uma análise quantitativa das atitudes ou opiniões dos participantes [FERAUCHE; SILVA; ITO, 2024].

3.5.3.3 Identificação das Medidas da Avaliação do Protocolo

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 [FERAUCHE; SILVA; ITO, 2024].

Com três etapas finalizadas, continua com a última etapa da elaboração do protocolo.

3.5.3.4 Definição de Níveis de Pontuação e Critérios de Julgamento do Protocolo

Para o protocolo QRAMHe, optou-se para o cálculo final pela média aritmética da pontuação de cada categoria. Definindo NQC como Nível de Qualidade por Categoria [FERAUCHE; SILVA; ITO, 2024].

A partir dos resultados obtidos para cada categoria, calcula-se o resultado que representa o nível de qualidade geral do aplicativo, pelo cálculo da média aritmética. Considerando NQCA como o Nível de Qualidade do Aplicativo [FERAUCHE; SILVA; ITO, 2024].

E, por fim, para definir o nível de qualidade, 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%. A Tabela 1 demonstra o nível de qualidade conforme o resultado obtido da pontuação por categoria.

Tabela 3.3 - Pontuações definidas para as respostas por média aritmética por categoria

Média Aritmética por Categoria	Significado da resposta	Nível de Qualidade
5	Eficiente	Superior
4	Aceitável	
3	Regular	Médio
2	Deficiente	Inferior
1	Ineficiente	

De acordo com os possíveis níveis de qualidade do aplicativo, considera-se que um aplicativo com nível superior possui um bom nível de qualidade, já que para obter essa pontuação o aplicativo conseguiu satisfazer os critérios de qualidade avaliados. Por outro lado, um nível de qualidade médio revela que o aplicativo necessita de melhorias. 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 deva passar por uma reestruturação, com base nos critérios de qualidade não satisfeitos a partir da avaliação conduzida [FERAUCHE; SILVA; ITO, 2024].

3.5.3.5 Validação do protocolo

Para avaliar o protocolo, os autores escolheram por dois processos de validação:

a. Primeira Avaliação

Para a primeira avaliação, estrutura do protocolo (a), opta-se pela aplicação de entrevista com especialistas. A entrevista é uma técnica usada em pesquisas qualitativas para obter credibilidade, precisão e a confiabilidade dos dados coletados [SIMPSON; MCDOWELL, 2019].

As etapas da entrevista de pesquisa envolvem planejamento, gerenciamento, análise e representação das entrevistas. Para o planejamento da entrevista, adaptando Mann (2016), são necessários os seguintes passos: (aa) objetivo; (ab) participantes; (ac) preparação; (ad) realização; e (ae) pós-entrevista.

Para o passo ‘aa’ do planejamento da entrevista, o objetivo da entrevista é validar o protocolo. No passo ‘ab’, para a seleção dos participantes é necessário profissionais qualificados que possam contribuir no processo de qualificação do protocolo [CHANG; MORREALE; BALON, 2017]. Para a preparação das entrevistas, passo ‘ac’, é necessário o prévio agendamento com previsão da duração e ambiente de condução. Na realização da entrevista, passo ‘ad’, as entrevistas devem ser conduzidas conforme a plataforma escolhida, respeitando o tempo agendado. No passo referente a pós-entrevista, passo ‘ae’, após o andamento da entrevista, ocorre a análise qualitativa e quantitativa. Para a análise qualitativa, é feita a transcrição, tradução e análise temática do discurso [BRITTEN, 1995]. A estatística Kappa de Fleiss é a medida estatística usada para avaliar a confiabilidade da concordância entre vários avaliadores [GWET, 2021]. Portanto, com o auxílio desta técnica é possível avaliar a reprodutibilidade do questionário.

Neste primeiro momento da análise, o resultado das entrevistas foi convertido em tabela classificando e resumindo os pareceres dos entrevistados, assim composta:

Tabela 3.4 - Conversão dos dados das entrevistas para análise quantitativa

Entrevistado	Questão 1	Questão 2	Questão 3	Questão 4	Questão 5	Questão 6	Questão 7	Questão 8	Questão 9	Questão 10
A	Sim	Parc.	Não	Sim						
B	Não	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Sim
C	Sim	Sim	Sim	Sim	Sim	Parc.	Sim	Não	Sim	Sim

Para a análise, foi utilizado o software IBM SPSS Statistics para Windows (versão 28.0) com nível de significância adotado 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]; $z = 2,032$; $p < 0,043$) [FERAUCHE; SILVA; ITO, 2024].

Ao analisar a categoria de respostas (SIM, NÃO, PARCIALMENTE), houve uma concordância significativa entre os entrevistados para a resposta SIM, representando uma confiabilidade/reprodutibilidade moderada ($k = 0,441$ [IC 95%: 0,083-0,799]; $z = 2,415$; $p < 0,017$), possibilitando a continuação do estudo. Diante destes resultados, seguiu-se para o segundo momento da análise que, após a seleção do aplicativo, o instrumento foi direcionado aos profissionais da área para aplicação do instrumento [FERAUCHE; SILVA; ITO, 2024].

Após a análise detalhada da estrutura do protocolo, foram realizados ajustes com base no feedback fornecido pelos entrevistadores [FERAUCHE; SILVA; ITO, 2024].

b. Segunda Avaliação

Para a segunda avaliação, para verificar a eficácia do protocolo, segue-se os seguintes procedimentos:

- **seleção do aplicativo:** para a seleção do aplicativo foi o Protocolo de Mapeamento Sistemático para Busca de Aplicativos de Saúde em Repositórios Não Acadêmicos (PBAS) escrito por Lima e Peres. Este protocolo faz uma adaptação do protocolo de mapeamento sistemático da literatura para o levantamento de aplicativos de software de saúde em repositórios não acadêmicos de aplicativos e código-fonte [LIMA; PERES, 2021].
- **aplicação da ferramenta QRAMHe no aplicativo selecionado:** foi utilizada a ferramenta QRAMHe em formato online (<https://qramhe.org/>), em um período determinado com uma seleção de participantes qualificados para contribuir no processo de avaliação de eficácia do protocolo [FERAUCHE; SILVA; ITO, 2024].
- **análise da aplicação:** fez-se a utilização das análise de coeficiente Alfa de Cronbach e o teste de Kappa de Fleiss [FERAUCHE; SILVA; ITO, 2024].

Foi utilizado o coeficiente Alfa de Cronbach, medida estatística que avalia a consistência e confiabilidade interna de uma escala, questionário ou instrumento de avaliação composto por múltiplas questões de um constructo. Para esse fim, o coeficiente Alfa de Cronbach varia de 0 a 1, indicando que os valores mais próximos de 1 evidenciam maior a consistência interna dos itens do questionário, ou seja, mede o quanto as questões referem-se ao mesmo assunto a que se propõe o construto e se são consistentes entre si [BLAND; ALTMAN, 1997] [GLIEM; GLIEM; OTHERS, 2003] [GASPAR; SHIMOYA, 2017] [FERAUCHE; SILVA; ITO, 2024]

O nível de consistência interna é considerado substancial para os valores acima de 0,60, e para valores acima de 0,80, quase perfeita [LANDIS; KOCH, 1977].

A análise desta etapa partiu das respostas de 10 profissionais da área de Saúde, especificamente, Educação Física, tendo como coeficiente Alfa de Cronbach significativo (Cronbach = 0,97; $p < 0,001$), evidenciando uma consistência interna e confiabilidade muito alta [FERAUCHE; SILVA; ITO, 2024].

3.5.4 Lições Aprendidas e Melhores Práticas

Diversas lições foram aprendidas no desenvolvimento do protocolo, alguns pontos podem ser destacados:

- Identificar quais os métodos de avaliações existentes para aplicativos na área da saúde é algo trabalhoso
- A colaboração entre profissionais de saúde, desenvolvedores de software e especialistas em avaliação foi essencial para garantir que o protocolo aborde todas as necessidades e especificidades dos usuários, pacientes e profissionais;
- A necessidade de adaptar o protocolo a diferentes tipos de aplicativos de saúde e contextos de uso foi uma lição fundamental. O protocolo deve ser suficientemente flexível para acomodar inovações tecnológicas e mudanças nas práticas de saúde;
- A validação contínua do protocolo permitiu ajustes e refinamentos antes da implementação final, garantindo maior precisão e relevância nas avaliações;
- fornecer treinamento adequado aos avaliadores sobre como aplicar o protocolo e interpretar os resultados identifica que o protocolo não foi suficiente independente para assegurar a consistência e a validade das avaliações;
- não se limitar à utilização de uma única escala para a avaliação, mas sim empregar diversas escalas conforme necessário, incluindo opções de seleção única e múltipla, dependendo da natureza de cada questionamento;
- cada um dos questionamentos pode ter valores diferenciados, permitindo que alguns aspectos sejam ponderados mais fortemente enquanto outros recebem menor peso, dependendo da importância relativa de cada critério na avaliação geral.

Apesar das lições compreendidas, algumas práticas foram identificadas:

- estabelecer critérios claros e específicos para cada dimensão de avaliação, como usabilidade, segurança e confiabilidade, facilita a aplicação consistente do protocolo e melhora a comparabilidade entre diferentes aplicativos;
- basear o desenvolvimento do protocolo em normas internacionais como a ISO 25000 proporcionou uma estrutura sólida e reconhecida globalmente, assegurando que o protocolo estivesse alinhado com as melhores práticas de qualidade de software;

- desenvolver questionários estruturado ajudou a sistematizar a coleta de dados e garantir que todos os aspectos relevantes fossem considerados durante a avaliação.

Incorporar essas lições aprendidas e melhores práticas pode não apenas fortalecer a aplicação do protocolo, mas também contribuir para o avanço da qualidade e segurança dos aplicativos de saúde em geral.

3.6 Desafios e Perspectivas da avaliação de Tecnologia em Saúde Digital

A avaliação de tecnologias digitais em saúde é um tema amplamente discutido em diversos estudos. O desenvolvimento dessas tecnologias ocorre em um ritmo bastante acelerado, o que torna crucial que os processos de avaliação sejam igualmente ágeis. Um dos desafios é como acelerar o desenvolvimento de processos e protocolos para a avaliação dessas tecnologias digitais diante de seu ritmo acelerado de desenvolvimento, de forma que a população seja beneficiada pela adoção dessas tecnologias.

Além disso, conforme destacado no estudo de MEZEI et al. [2023], há uma necessidade de padronização para a classificação, avaliação e financiamento dessas tecnologias. A falta de padrões dificulta a comparação entre tecnologias e a tomada de decisões informadas por parte dos stakeholders.

As tecnologias digitais aplicadas na saúde abrangem uma vasta gama, incluindo aplicativos móveis, algoritmos de inteligência artificial e aprendizado de máquina, robótica e Internet das Coisas médicas. Essa variedade dificulta o desenvolvimento de uma única ferramenta de avaliação que possa ser aplicada a todas essas tecnologias, como evidenciado pelas diversas propostas de frameworks existentes. A diversidade dessas tecnologias requer abordagens específicas e adaptadas para cada tipo de inovação.

Ainda conforme mencionado no estudo de MEZEI et al. [2023], vários países têm desenvolvido seus próprios frameworks para a avaliação de tecnologias digitais. Embora esses frameworks tenham dimensões similares, eles possuem diferenças que refletem as necessidades e contextos específicos de cada país. Isso sugere que, enquanto a padronização é desejável, a flexibilidade para acomodar variabilidades locais também é necessária.

Mesmo quando focados em tecnologias específicas, como aplicativos móveis de saúde, existem várias propostas de protocolos de avaliação. Esses protocolos, porém, não demonstram uniformidade nas dimensões avaliadas, evidenciando a necessidade de padronização e especificidade nos processos de avaliação.

A diversidade dos quadros de Avaliação de Tecnologias em Saúde (ATS) para Tecnologias de Saúde Digital (TDS) decorre da complexidade da avaliação dessas tecnologias em rápida evolução, que utilizam métodos diversos e transformam significativamente os sistemas e comportamentos sociais e de saúde. A rápida evolução das TDSs exige métodos de avaliação dinâmicos e adaptáveis que possam acompanhar essas mudanças.

Nas propostas de frameworks de avaliação de tecnologias em saúde, é comum encontrar uma dimensão dedicada à cibersegurança dos dados de saúde. No entanto, não é explicitamente observada uma dimensão dedicada à segurança do paciente ao avaliar essas tecnologias digitais. A segurança do paciente é uma preocupação crítica que deve ser integrada nos processos de avaliação para garantir que as tecnologias digitais não

apenas protejam os dados, mas também melhorem a segurança e a qualidade do cuidado ao paciente.

References

ABIODUN, T.; RAMPERSAD, G.; BRINKWORTH, R. Driving Industrial Digital Transformation. **Journal of Computer Information Systems**, v. 63, n. 6, p. 1345–1361, 2 nov. 2023.

ABNT NBR ISO/IEC 25062. **ABNT NBR ISO/IEC 25062 NBRISO/IEC25062 Engenharia**. Disponível em: <<https://www.normas.com.br/visualizar/abnt-nbr-nm/30823/abnt-nbriso-iec25062-engenharia-de-software-requisitos-e-avaliacao-da-qualidade-de-produto-de-software-square-formato-comum-da-industria-fci-para-relatorios-de-teste-de-usabilidade>>. Acesso em: 28 maio. 2024.

ABRAN, A. Software Project Estimation: The Fundamentals for Providing High Quality Information to Decision Makers. 6 abr. 2015.

AGARWAL, S. et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. **BMJ**, v. 352, p. i1174, 17 mar. 2016.

AINI, Q.; ZULIANA, S. R.; SANTOSO, N. P. L. Management Measurement Scale As A Reference To Determine Interval In A Variable. **Aptisi Transactions on Management (ATM)**, v. 2, n. 1, p. 45–54, 2 jan. 2018.

AKBAR, S.; COIERA, E.; MAGRABI, F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. **Journal of the American Medical Informatics Association**, v. 27, n. 2, p. 330–340, 1 fev. 2020.

ALEXANDRE, N. M. C.; COLUCI, M. Z. O. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. **Ciência & Saúde Coletiva**, v. 16, n. 7, p. 3061–3068, jul. 2011.

AZAD-KHANEGHAH, P.; RODUTA ROBERTS, M.; LIU, L. Alberta Rating Index for Apps: Study of Reliability and Validity. **Canadian Journal of Occupational Therapy**, v. 89, n. 3, p. 326–338, 1 set. 2022.

BARUA, A. Methods for Decision-Making in Survey Questionnaires Based on Likert Scale. **Journal of Asian Scientific Research**, v. 3, n. 1, p. 35–38, 1 jan. 2013.

BARZEGAR, A. **Measuring Software Quality Product Based on Fuzzy Inference System Techniques in ISO Standard**. Preprints, , 27 abr. 2021. Disponível em: <<https://www.preprints.org/manuscript/202104.0721/v1>>. Acesso em: 9 maio. 2024

BOUGROUN, Z. et al. **A Survey for Validation Concepts to Measure Quality as Well Their Application on the Maintainability of ISO.** (A. El Moussati et al., Eds.) *Advances in Smart Technologies Applications and Case Studies. Anais...* Cham: Springer International Publishing, 2020.

BRUCE, C. et al. Evaluating Patient-Centered Mobile Health Technologies: Definitions, Methodologies, and Outcomes. **JMIR mHealth and uHealth**, v. 8, n. 11, p. e17577, 11 nov. 2020.

BUCHNER, B. Datenschutz und Datensicherheit - DuD. **DIGITAL HEALTH**, v. 46, n. 12, p. 729–729, 1 dez. 2022.

DALATI, S. Measurement and Measurement Scales. Em: MARX GÓMEZ, J.; MOUSELLI, S. (Eds.). **Modernizing the Academic Teaching and Research Environment: Methodologies and Cases in Business Research.** Cham: Springer International Publishing, 2018. p. 79–96.

ELIAS, F. T. S. A importância da avaliação de tecnologias para o Sistema Único de Saúde. 2013.

ENTRINGER, T. C.; FERREIRA, A. DA S. Software quality evaluation: a bibliometric analysis and future perspectives. **Independent Journal of Management & Production**, v. 10, n. 5, p. 1499–1515, 1 out. 2019.

EUNETHTA. **EUnethTA Joint Action 2 Work Package 8.** , 2016. Disponível em: <<https://www.eunethta.eu/wp-content/uploads/2018/03/HTACoreModel3.0-1.pdf>>. Acesso em: 14 maio. 2024

EUROPEAN COMMISSION. DIRECTORATE GENERAL FOR HEALTH AND FOOD SAFETY. Opinion on assessing the impact of digital transformation of health services. **LU: Publications Office**, 2019.

FERAUCHE, T. M. Y.; SILVA, J. J.; ITO, M. Proposal for the Protocol for Assessment of mHealth: QRAMHe. **CONTECSI - International Conference on Information Systems and Technology Management**, n. 20, 2024.

FRUHLING, A.; LEE, S. Assessing the Reliability, Validity and Adaptability of PSSUQ. p. 10, 2005.

FUKUYAMA, M. Society 5.0: Aiming for a new human-centered society. **Japan Spotlight**, v. 27, n. 5, p. 47–50, 2018.

GABINETE DO MINISTRO. **Portaria n. 2.510.** Disponível em: <https://bvsms.saude.gov.br/bvs/saudelegis/gm/2005/prt2510_19_12_2005.html>.

GARVIN, D. A. **Gerenciando A Qualidade**. 3ª edição ed. [s.l.] Qualitymark, 2002.

GIACOMETTI, A.; MARKHOFF, B.; SOULET, A. Comparison Table Generation from Knowledge Bases. Em: VERBORGH, R. et al. (Eds.). **The Semantic Web**. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021. v. 12731p. 179–194.

GOODMAN, C. S. **Introduction to health care technology assessment: ten basic steps**. , 1998. Disponível em: <<http://www.nlm.nih.gov/nichsr/ta101>>

IMDRF, S. AS A M. D. (SAM) W. G. “Software as a Medical Device”: Possible Framework for Risk Categorization and Corresponding Considerations. **International Medical Device Regulators Forum (IMDRF)**, 2014.

ISO 9126. **ISO/IEC 9126-1:2001**. Disponível em: <<https://www.iso.org/standard/22749.html>>. Acesso em: 15 out. 2023.

ISO 14598. **ISO/IEC 14598-5:1998**. Disponível em: <<https://www.iso.org/standard/24906.html>>. Acesso em: 15 out. 2023.

ISO 25000. **ISO/IEC 25000:2014**. Disponível em: <<https://www.iso.org/standard/64764.html>>. Acesso em: 9 maio. 2024.

ISO 82304-2. **ISO/TS 82304-2:2021 Health software — Part 2: Health and wellness apps — Quality and reliability**. Disponível em: <<https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/81/78182.html>>. Acesso em: 16 set. 2022.

ITO, M. Patient-centered care. Em: **Fundamentals of Telemedicine and Telehealth**. [s.l.] Elsevier, 2020. p. 115–126.

ITO, M.; PEREIRA, D. **Proposta de um Ambiente Colaborativo para a Coordenação de Cuidado de Pacientes Crônicos**. Anais Principais do Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS 2015). **Anais...** Em: NCIPAI DO SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA À SAÚDE. Brasil: Sociedade Brasileira de Computação - SBC, 20 jul. 2015. Disponível em: <<https://sol.sbc.org.br/index.php/sbcas/article/view/10382>>. Acesso em: 26 maio. 2024

JUNG, H.-W. Validating the external quality subcharacteristics of software products according to ISO/IEC 9126. **Computer Standards & Interfaces**, v. 29, n. 6, p. 653–661, 1 set. 2007.

KAPUR, R. et al. **Training Software Engineers for Qualitative Evaluation of Software Architecture**. Disponível em: <<https://arxiv.org/abs/2105.09595v1>>. Acesso em: 9 maio. 2024.

KARUNA, P.; DIVYA, M.; MANGALA, N. Statistical Analysis of Metrics for Software Quality Improvement. **International Journal of Software Engineering & Applications**, v. 9, n. 1, p. 77–88, 30 jan. 2018.

KEENEY, S.; HASSON, F.; MCKENNA, H. The Delphi Technique in Nursing and Health Research. Em: **The Delphi Technique in Nursing and Health Research**. [s.l.] John Wiley & Sons, Ltd, 2011. p. 1–17.

KOTTNER, J. et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. **International Journal of Nursing Studies**, v. 48, n. 6, p. 661–671, jun. 2011.

KUI, K. M.; ALI, K. B.; SURYN, W. The Analysis and Proposed Modifications to ISO/IEC 25030—Software Engineering—Software Quality Requirements and Evaluation—Quality Requirements. **Journal of Software Engineering and Applications**, v. 9, n. 4, p. 112–127, 18 abr. 2016.

LANDIS, J. R.; KOCH, G. G. The Measurement of Observer Agreement for Categorical Data. **Biometrics**, v. 33, n. 1, p. 159–174, 1977.

LESQUELEN, A. et al. OP124 Can Registry Failures Be Compensated By Medico-Administrative Database. **International Journal of Technology Assessment in Health Care**, v. 33, n. S1, p. 56–57, jan. 2017.

LIAROPOULOS, L. Do we need ‘care’ in technology assessment in health care, letter to the editor. **International Journal of Technology Assessment in Health Care**, v. 13, n. 1, p. 125–127, 1997.

LIMA JUNIOR, F.; MARTIMIANO, L. Avaliação da qualidade de softwares voltados à gestão de projetos. **XXX Encontro Nacional de Engenharia de Produção, São Carlos. Anais do XXX Encontro Nacional de Engenharia de Produção**, 2010.

LIMA, L. F. D.; PERES, L. M. **Protocolo de Mapeamento Sistemático para Busca de Aplicativos de Saúde em Repositórios Não-acadêmicos**. Anais do I Workshop de Práticas de Ciência Aberta para Engenharia de Software (OpenScienSE 2021). **Anais...** Em: WORKSHOP DE PRÁTICAS DE CIÊNCIA ABERTA PARA ENGENHARIA DE SOFTWARE. Brasil: Sociedade Brasileira de Computação, 28 set. 2021. Disponível em: <<https://sol.sbc.org.br/index.php/opensciense/article/view/17138>>. Acesso em: 10 nov. 2023

LU, Y. et al. **A Quality Evaluation Method for Software Testing about Safety-Critical Software**. 2022 IEEE 13th International Conference on Software Engineering and Service Science (ICSESS). **Anais...** Em: 2022 IEEE 13TH INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING AND SERVICE SCIENCE (ICSESS). out. 2022. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/9930258>>. Acesso em: 9 maio. 2024

MAASS, L. et al. The Definitions of Health Apps and Medical Apps From the Perspective of Public Health and Law: Qualitative Analysis of an Interdisciplinary Literature Overview. **JMIR mHealth and uHealth**, v. 10, n. 10, p. e37980, 31 out. 2022.

MARSHALL, G. The purpose, design and administration of a questionnaire for data collection. **Radiography**, v. 11, n. 2, p. 131–136, maio 2005.

MAXIM, B.; PRESSMAN, R. S. **Engenharia de software: uma abordagem profissional**. Porto Alegre:[sn], , 2021.

MCHUGH, M. L. Interrater reliability: the kappa statistic. **Biochemia Medica**, v. 22, n. 3, p. 276–282, 15 out. 2012.

MESKÓ, B. et al. Digital health is a cultural transformation of traditional healthcare - PMC. 2017.

MEZEI, F. et al. International practices in health technology assessment and public financing of digital health technologies: recommendations for Hungary. **Frontiers in Public Health**, v. 11, p. 1197949, 31 ago. 2023.

MHIMDI, N.; KARAA, W. B. A.; GHEZALA, H. B. Data Collection in Hospital Emergencies: A Questionnaire Survey. **International Journal of Computer and Information Engineering**, v. 16, n. 3, p. 48–53, 7 mar. 2022.

MINGLIANG, L. et al. **Research on the Evaluation Method of Software Static Quality based on AHP**. Proceedings of the 2021 1st International Conference on Control and Intelligent Robotics. **Anais...**: ICCIR '21. New York, NY, USA: Association for Computing Machinery, 13 ago. 2021. Disponível em: <<https://doi.org/10.1145/3473714.3473836>>. Acesso em: 9 maio. 2024

MINISTÉRIO DA SAÚDE. Resolução da Diretoria Colegiada (RDC) 657/2022 da Anvisa. . 24 mar. 2022 a.

MINISTÉRIO DA SAÚDE. Resolução da Diretoria Colegiada (RDC) 751/2022 da Anvisa. . 15 set. 2022 b.

MINOV, J. Evaluation of software applications. 14 dez. 2012.

OBAIDAT, M. S.; BOUDRIGA, N. Fundamentals of Performance Evaluation of Computer and Telecommunication Systems. 26 jan. 2010.

OMS. World Health Organisation. . 2012.

PASSOS, M. F. D. Elaboração e validação de escala de diferencial semântico para avaliação de personalidade. 27 abr. 2015.

PERES, L. DA S.; BUENO, A. F.; PIASSON, D. **MATURIDADE LOGÍSTICA DE PSLs: AVALIAÇÃO DA ADERÊNCIA DAS ESCALAS DE LIKERT E GUTTMAN EM UM MODELO ESTAGIADO XXXVI ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO.** [s.l: s.n.].

POLIT, D. F.; BECK, C. T. The content validity index: Are you sure you know what's being reported? critique and recommendations. **Research in Nursing & Health**, v. 29, n. 5, p. 489–497, out. 2006.

PRESIDÊNCIA DA REPÚBLICA - CASA CIVIL. Lei 12.401 de 2011. **Diário Oficial da União**, 28 abr. 2011.

ROBERTS, A. E. et al. 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**, v. 24, p. 100379, 1 abr. 2021.

ROOS, C.; SARTORI, S.; GODOY, L. P. Portaria n. 2.510. **Revista Produção Online**, v. 9, n. 3, 17 ago. 2009.

SANTOS, C. M. DOS et al. A utilização da escala de Borg em atividade física com gestantes. **ConScientiae Saúde**, v. 13, n. 2, p. 241–245, 14 jul. 2014.

SAÚDE 4.0. **PROPOSTAS PARA IMPULSIONAR O CICLO DAS INOVAÇÕES EM DISPOSITIVOS MÉDICOS (DMAs) NO BRASIL.** Saúde 4.0, , 2015.

SHAFFER, F. A.; PREZIOSI, P. Digital Health: Join the Movement. **Nursing Administration Quarterly**, v. 43, n. 4, p. 351–353, out. 2019.

SHEPHERD, J. What is the Digital Era? **Social and Economic Transformation in the Digital Era**, 2004.

SILVA, S. V. et al. **Avaliação da qualidade de um sistema acadêmico: estudo de caso no Q-Acadêmico.** Congresso de Tecnologia da Informação. **Anais...**2015.

SIMON LIN LINWOOD, M. D. Preface. Exon Publications. p. ix–x, 29 abr. 2022.

STEVENS, S. S. On the Theory of Scales of Measurement. **Science**, v. 103, n. 2684, p. 677–680, 7 jun. 1946.

STOUMPOS, A. I.; KITSIOS, F.; TALIAS, M. A. Digital Transformation in Healthcare: Technology Acceptance and Its Applications. **International Journal of Environmental Research and Public Health**, v. 20, n. 4, p. 3407, 15 fev. 2023.

STOYANOV, S. R. et al. Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps. **JMIR mHealth and uHealth**, v. 3, n. 1, p. e3422, 11 mar. 2015.

STOYANOV, S. R. et al. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). **JMIR mHealth and uHealth**, v. 4, n. 2, p. e5849, 10 jun. 2016.

TRENDOWICZ, A.; KOPCZYŃSKA, S. Chapter 4 - Adapting Multi-Criteria Decision Analysis for Assessing the Quality of Software Products. Current Approaches and Future Perspectives. Em: MEMON, A. (Ed.). **Advances in Computers**. Advances in Computers. [s.l.] Elsevier, 2014. v. 93p. 153–226.

VON HUBEN, A. et al. Health technology assessment for digital technologies that manage chronic disease: a systematic review. **International Journal of Technology Assessment in Health Care**, v. 37, n. 1, p. e66, 2021.

VUONG, Q.-H. et al. The psychological mechanism of internet information processing for post-treatment evaluation. **Heliyon**, v. 8, n. 5, p. e09351, maio 2022.

WETTER, T. **Consumer health informatics**. [s.l.] Springer, 2012.

WHO, W. H. O. Global strategy on digital health 2020–2025. 2021. **License: CC BY-NC-SA**, v. 3, 2021.

YADAV, K. K. Software Testing and Evaluation. 1 jan. 2015.

YAMAKOSHI, K.; NISHIYAMA, M. Software evaluation method and software evaluation device. 27 set. 2018.

YANG, H. Measuring Software Product Quality with ISO Standards Base on Fuzzy Logic Technique. Em: LUO, J. (Ed.). **Affective Computing and Intelligent Interaction**. Berlin, Heidelberg: Springer, 2012. p. 59–67.

YUSOFF, M. S. B. ABC of Content Validation and Content Validity Index Calculation. **Education in Medicine Journal**, v. 11, p. 49–54, 28 jun. 2019.

ZARILI, T. F. T. et al. Técnica Delphi no processo de validação do Questionário de Avaliação da Atenção Básica (QualiAB) para aplicação nacional. **Saúde e Sociedade**, v. 30, p. e190505, 18 jun. 2021.

ZHANG, Y. et al. **Development and Reconstitution of Software Quality Measurement and Evaluation Standards**. 2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). **Anais...** Em: 2018 19TH IEEE/ACIS INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING, ARTIFICIAL INTELLIGENCE, NETWORKING AND PARALLEL/DISTRIBUTED COMPUTING (SNPD). jun. 2018. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/8441040>>. Acesso em: 9 maio. 2024

ZHOU, L. et al. The mHealth App Usability Questionnaire (MAUQ): Development and Validation Study. **JMIR mHealth and uHealth**, v. 7, n. 4, p. e11500, 11 abr. 2019.