UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

João Marcus Alves

THE ADEQUATE MODEL: SOFTWARE QUALITY EVALUATION MODEL FOR TELEMEDICINE AND TELEHEALTH SYSTEMS

Florianópolis

2017

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Alves, João Marcus
The AdEQUATE Model : Software quality evaluation model for telemedicine and telehealth systems / João Marcus Alves ; orientadora, Christiane Gresse von Wangenheim, coorientador, Alexandre Savaris, 2017.
139 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós Graduação em Ciência da Computação, Florianópolis, 2017.

Inclui referências.

1. Ciência da Computação. 2. Software quality. 3. Software quality evaluation. 4. Software engineering. I. von Wangenheim, Christiane Gresse. II. Savaris, Alexandre. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Ciência da Computação. IV. Título.

João Marcus Alves

THE ADEQUATE MODEL: SOFTWARE QUALITY EVALUATION MODEL FOR TELEMEDICINE AND TELEHEALTH SYSTEMS

Esta dissertação foi julgada adequada para obtenção do título de mestre e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

Florianópolis, ____de __________________ de _________.

____
Prof. José Luís Almada Güntzel, Dr.
Coordenador do Programa

____
Prof. Alexandre Savaris, Dr.
Universidade Federal de Santa Catarina
Coorientador

Banca Examinadora:

____
Prof[a]. Doc. rer. Nat. Christiane Gresse von Wangenheim, PMP
Universidade Federal de Santa Catarina
Orientadora

____
Prof. Luiz Ary Messina, Dr.
Instituto de Informática e Automação (videoconferência)

____
Prof. Adriano Ferreti Borgatto, Dr.
Universidade Federal de Santa Catarina

____
Prof[a]. Patrícia Della Méa Plentz, Dra.
Universidade Federal de Santa Catarina

To Marcus, Joaquina and Justiniana, in memoriam.

ACKNOWLEDGMENTS

To Professors Christiane von Wangenheim and Alexandre Savaris for guiding and advising me throughout those years.

To my precious family, my fiance Raquel, my mother Jadet, my father Frederico, my sister Alessandra, my brother-in-law Alexandre, my uncle João, my grandmother Joaquina and grandfather Avelino. This work could not have been done without their love and companionship.

To all colleagues and friends of LabTelemed and INCoD: Andrei, Eduardo, Frederico, Gabi, Jean, Mila, Minas, Nathalie, Paulo, Rodrigo, Vinícius, Walter. For the kindest companionship and cups of coffee when they are needed the most.

To all STT/DATATOX/LACEN users who dedicated their precious time answering the AdEQUATE questionnaire, for their time and feedback.

To the Federal University of Santa Catarina.

To the members of the evaluation board.

Last but not least I would thank all my cats for their undying affection: Lydia, Sauron, Satan, Dracula, and Frygga.

All that is gold does not glitter,
Not all those who wander are lost;
The old that is strong does not wither,
Deep roots are not reached by the frost.
From the ashes, a fire shall be woken,
A light from the shadows shall spring;
Renewed shall be blade that was broken,
The crownless again shall be king

J.R.R. Tolkien, 1954

RESUMO

Sistemas de *software* usado para o fornecimento de saúde deveriam satisfazer as necessidades explícitas e implícitas de seus vários *stakeholders*, em outras palavras, ter um alto grau de qualidade de *software* para reduzir, por exemplo, erros em receitas e diagnósticos médicos, para evitar efeitos negativos na saúde de pacientes. Nesse contexto, modelos de qualidade de *software* são desenvolvidos para medir o quanto um sistema de *software* está operando com qualidade. Esses modelos decompoem o conceito de qualidade em características e subcaracterísticas mais compreensíveis. Contudo, não é trivial avaliar a percepção de qualidade de *software* de usuários finais em sistems de saúde baseados na Tecnologia de Informação e Comunicação (TIC) – um termo guarda-chuva que engloba vários domínios de fornecimento remoto de saúde, como telemedicina (uso de telecomunicações para diagnóstico médico e tratamento de paciente) e telessaúde (fomento de saúde e prevenção de doenças usando TIC). Assim, essa dissertação de mestrado apresenta um modelo de avaliação de qualidade de *software* para sistemas de telemedicina e telessaúde. Para desenvolver este modelo, foi realizado uma revisão sistemática da literatura em modelos de qualidade de *software* que focam especificamente na qualidade de *software* na percepção do usuário final. Uma vez definido o estado da arte em tais modelos, usando o paradígma Objetivo/Questão/Métrica, foi desenvolvido um modelo de qualidade de *software* juntamente com seu instrumento de medição. O modelo foi usado para avaliar a qualidade de *software* de 6 sistemas de telemedicina e telessaúde usados no contexto Brasileiro de saúde pública, e os dados obtidos nesse estágio da pesquisa foram usados para validar o modelos em termos de validade de confiabilidade e construção. Essa pesquisa definiu o estado da arte em modelos de qualidade de *software* no contexto de sistemas de saúde baseados em TIC, e desenvolveu um modelo de qualidade de *software* juntamente com um instrumento de medição que pode ser usado para avaliar qualidade de *software* em diferentes sistemas de telemedicina e telessaúde em diferentes contextos de uso por seus usuários finais um questionário de 68 items. Essa pesquisa, dentro de seu contexto, usou o modelo proposto para avaliar a qualidade de *software* de múltiplos sistemas de telemedicina e telessaúde, assim obtendo um conjunto de dados de respostas de usuários finais sobre qualidade de *software* no contexto de sistemas de telemedicina e telessaúde que pode ser usado

para fomentar pesquisas futuras no tópico; e levantou as evidências de validade de confiabilidade e construção do modelo e seu instrumento de medição.
Palavras-chave: Qualidade de *software*. Modelo de Qualdiade de *software*. Avaliação de Qualidade de *software*. Telemedicina. Telessaúde.

RESUMO EXPANDIDO

Introdução

Atualmente é impossível praticar a medicina moderna sem Tecnologias de Informação (TI) (SHORTLIFFE; CIMINO, 2013), e nesse contexto, o conceito de saúde baseado em Tecnologias de Informação e Comunicação (TIC) é definido como a intersecção de informática médica, saúde pública, negócios no que se refere a serviços e informações de saúde entregues ou melhoradas usando a Internet e tecnologias correlatas (EYSENBACH, 2001). Tais técnicas se mostraram melhorar o acesso à saúde em comunidades rurais e remotas (SMITH; GRAY, 2009) e particularmente em paciente com doenças agudas ou crônicas, porque melhora a vigilância, diminui o custo e diminui os erros em medicamentos (CHAUDRY et al., 2006). Nesse contexto, a engenharia de software usa de modelos de qualidade de software para melhorar a qualidade de sistemas. Contudo, a despeito da importância desses modelos, a avaliação de sistemas como telemedicina e telessaúde é raramente comentada (SÁNCHEZ-PI; MOLINA, 2010).

Embora algumas pesquisas abordem a qualidade de sistemas de saúde baseados em TIC, a maioria deles é relacionado a uma aplicação específica (LEROUGE; GARFIELD; HEVNER, 2002; LEROUSE et al., 2004), porque avaliar sistemas de saúde baseados em TIC não é trivial devido à natureza complexa do domínio da saúde (ANDARGOLI et al., 2017), e de não haver um consenso sobre como avaliar a qualidade de tais sistemas.

Objetivos

Este trabalho propõe e avalia um novo modelo para avaliação de qualidade de *software* para sistemas de telemedicina e telessaúde, o modelo AdEQUATE do ponto de vista dos usuários finais. O modelo proposto é decomposto em características de qualidade sistematicamente derivadas dos modelos da norma ISO/IEC 25010. Ainda no contexto dessa pesquisa deve-se desenvolver um instrumento de medição e guia de análise para a avaliação.

Metodologia

Esta pesquisa representa uma pesquisa aplicada com um objetivo exploratório e utiliza uma abordagem quantitativa e qualitativa, documental e procedimentos de estudos de caso. É uma pesquisa aplicada pois busca resolver o problema exposto (a lacuna no campo atual de modelos de avaliação de qualidade de *software*) (KOTHARI, 2004). Também pode ser considerada uma pesquisa qualitativa porque usa dados qualitativos coletados utilizando o instrumento de medição desenvolvido (KOTHARI, 2004). Da mesma maneira, é uma pesquisa exploratória pois pode levantar a hipótese das razões que causaram a percepção do grau de qualidade de *software* em contextos de sistemas de telemedicina e telessaúde. Para atingir os objetivos propostos, essa pesquisa usa uma abordagem multi-método, dividida em 4 estágios. O primeiro estágio se refere às técnicas e passos utilizados para a síntese do estado da arte; o segundo mostra os métodos utilizados no desenvolvimento do modelo; o terceiro estágio se refere aos métodos utilizados na execução dos estudos de caso; o último estágio discorre sobre os métodos utilizados na avaliação do modelo.

Resultados e Discussão

O resultado da avaliação do modelo AdEQUATE, baseado nos dados coletados nos estudos de caso, mostram que o modelo pode analisar a qualidade de software de sistemas de telemedicina e telessaúde, com confiabilidade e tem uma moderada validade de construção. Contudo, a análise fatorial revela que o modelo pode considerar itens diferentes no instrumento de medição, alterando a pertinência desses itens em suas respectivas características de qualidade, ou removendo alguns itens completamente.

Considerações Finais

Através de uma revisão sistemática da literatura, foi sintetizado o estado da arte em modelos de avaliação de qualidade de software em telemedicina e telessaúde. Baseado nessa síntese, um novo modelo de qualidade de software foi proposto, com uma decomposição similar à proposta pelos modelos do padrão ISO/IEC 25010. O questionário AdEQUATE desenvolvido como instrumento de medição foi avaliado e considerado apto a coletar a percepção de qualidade de software de sistemas de telemedicina e telessaúde. Este trabalho abre uma gama de pesquisas, incluindo a revisão do modelo segundo os resultados da análise fatorial.

Palavras-chave: Qualidade de *software*. Modelo de qualidade de software. Avaliação de qualidade de *software*. Telemedicina. Telessaúde.

ABSTRACT

Software systems used for healthcare delivery should satisfy the stated and implied needs of its various stakeholders, in other words, to have a high software quality degree to reduce, for instance, prescription or diagnose errors, and to avoid negative effects on a patient's health status. In this context, software quality models are developed to measure how a software system is operating with quality. These models decompose the concept of quality into more manageable characteristics and subcharacteristics. However it is not trivial to evaluate the perception of software quality by the end users in Information and Communication Technology (ICT) based healthcare systems – an umbrella term that encompasses a wide range of remote healthcare delivery domains, like telemedicine (the use of telecommunications for medical diagnosis and patient care) and telehealth (health promotion and disease prevention using ICT). Therefore, this master theses presents a software quality model for evaluating telemedicine and telehealth systems. To develop such model we have performed a systematic literature review on software quality models focusing specifically on the evaluation of software quality from the viewpoint of the end user. Once we have synthesized the state-of-the-art in such models, we, using Goal/Question/Metric paradigm, developed a software quality model along with a measuring instrument. The model was used to evaluate the software quality of 6 telemedicine and telehealth systems used in the context of Brazilian public health care, and data gathered during this stage in the research was used to validate the model in terms of reliability and construct validity. This research has synthesized the state-of-the-art in software quality evaluation models in the context of ICT-based healthcare systems, developed a software quality model that can be used to evaluate software quality of different telemedicine and telehealth systems in different contexts of use by its end users along with the model's measuring instrument – a 68 item questionnaire. We also have, within the scope of this research, used the proposed model to evaluate the software quality of multiple telemedicine and telehealth systems, thus gathering a data set of end-user responses about software quality in the context of telemedicine and telehealth systems that can be used to foster future research in the topic; and assessed the evidence of its reliability and construct validity.

Keywords: Software Quality. Software Quality Model. Software Qua-

lity Evaluation. Telemedicine. Telehealth.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

LIST OF SYMBOLS

TABLE OF CONTENTS

1 INTRODUCTION

In this chapter we present a brief introduction to software quality evaluation, defining the context of this work, discuss the problem detected and list the objectives of this research. This chapter also presents the scientific methodology used and shows the complete work structure.

1.1 CONTEXT

The first pandemic of the 21^{st} century has been acknowledged in 2009 (CHAN, 2009) requiring biochemical and clinical analysis of thousands of patients for diagnosis, treatment, and to control the pandemic. Today we know that combined efforts of multiple sectors have contributed to fighting the pandemic and post-pandemic. However, it is impossible to quickly and safely control a viral infection of such magnitude without ICT-based healthcare systems.

Nowadays is impossible to practice modern medicine without IT (SHORTLIFFE; CIMINO, 2013), and in this context, the concept of ICT-based healthcare, defines the intersection of medical informatics, public health, and business, referring to health services and information delivered or enhanced through the Internet and related technologies (EYSENBACH, 2001). This broad definition intentionally encompasses other concepts such as telemedicine (the use of telecommunications for medical diagnosis and patient care (MEA, 2001)) and telehealth (health promotion and disease prevention using ICT (KOCH, 2006; EYSENBACH, 2001)), besides m-health ("personalization" of ICT-based healthcare) and e-health (BASHSHUR et al., 2011). In this way, ICT-based healthcare encompasses applications that range from a system that aid in diagnosis (SINGH; GRABER, 2010) of diverse modalities (EKELAND; BOWES; FLOTTORP, 2010), electronic health records systems (GAI et al., 2015) to systems that facilitate medical education (WOOTTON, 2009), among others. Having such a variety of possible applications, ICT-based healthcare supports a number of different stakeholders, such as physicians, nurses, those responsible for the establishment of health care policies and procedures and researchers engaged in data collection (LEROUGE; GARFIELD; COLLINS, 2012).

ICT-based healthcare has been shown to improve the access to health care for people in rural and remote communities (SMITH; GRAY,

2009) – particularly for patients with prolonged acute or chronic illnesses, enhances disease surveillance, lowers the cost and decreases medication errors (CHAUDHRY et al., 2006). This suggests that ICT-based healthcare delivery has advantages over the the regular face-to-face practice of healthcare delivery (HAILEY; ROINE; OHINMAA, 2002), even more so in teleradiology (telemedicine for radiology speciality), telemental health (telemedicine for psychology treatment), transmission of echocardiographic images, teledermatology (telemedicine for the dermatology specialty), home telecare (telemedicine for monitoring patients at home) and on some medical consultations (HAILEY; ROINE; OHINMAA, 2002). For example, medication errors in hospitals are common, expensive, and sometimes harmful to patients (RADLEY et al., 2013). In this case, CPOE systems (a kind of ICT-based healthcare system that controls medications and prescriptions) may decrease the likelihood of medication error by 48% (RADLEY et al., 2013).

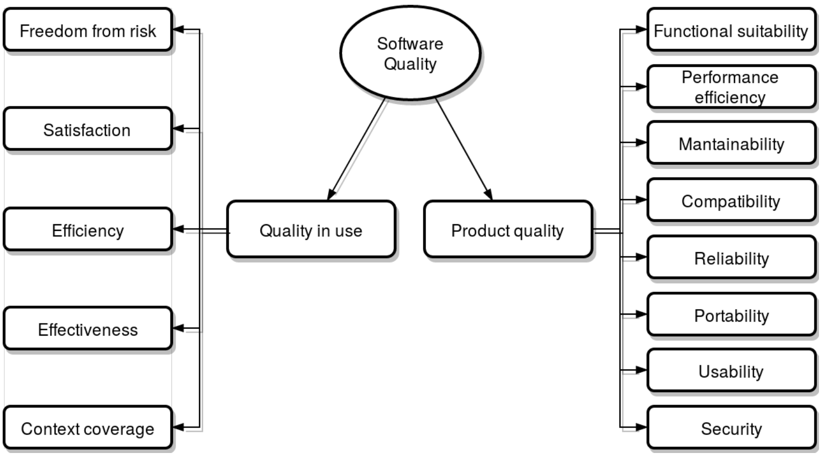
Regarding diagnose errors, the research of Singh et al. (2013), shows the severe impact of diagnosing errors in the patient's treatment. Furthermore, the same research shows that, from 190 cases of diagnosing errors, almost 4% were due to malfunctioning of the result system used by the hospital (SINGH et al., 2013).Furthermore, a Finish study that considers the opinion of over 3900 physicians shows that only 25% of them considers that the ICT system used helps to prevent medication errors, and 22% strongly disagree with that statement (VIITANEN et al., 2011). In this case, and in many others, the software quality of the providing ICT-based healthcare system affects the quality of healthcare offered, defined as the quality of health care offered is the effective convergence of technical and social factors using a collaborative technology used to produce the desired results (LEROUGE; HEVNER; COLLINS, 2007).

In this context, software engineering uses systematic software processes to develop high-quality software (FUGGETTA; NITTO, 2014), and in the ICT-healthcare domain, e.g., a software that mitigates diagnose or prescriptions errors. Thus, the role of software quality evaluation is each day more important (BREU; KUNTZMANN-COMBELLES; FELDERER, 2014). However, despite the evaluation of software quality in such systems be challenging and of critical importance, the issue is seldom tackled (SÁNCHEZ-PI; MOLINA, 2010).

Traditionally software quality, defined as the degree to which a system satisfies the stated and implied needs of its various stakeholders (International Organization for Standardization, 2014), is evaluated using software quality models. Likewise, quality models are defined as well-accepted means to support quality control of software systems, developed to measure how a software system is operating with quality (WAGNER, 2013). Often, software quality models decompose the abstract concept of software quality into quality characteristics (International Standard Organization, 2010; International Organization for Standardization, 2014; WAGNER, 2013) each one representing an aspect that should be present in a high-quality software system.

For example, the international standard ISO/IEC 25010 decomposes software quality into 13 characteristics such as usability, effectiveness, and security as shown in Fig. 1. Those characteristics may be further decomposed into sub characteristics (that again can be decomposed into sub-sub characteristics) (International Organization for Standardization, 2014).

Figure 1 – ISO/IEC 25010 decomposition of software quality



Fonte: Adapted from International Organization for Standardization (2014)

However, such software quality models are generic, being applicable to any software systems, but there are fundamental differences between different kinds of systems and application domains in terms of how they are used, which mean that different quality criteria apply (MOODY, 2005). This suggests that no single quality model may be able to serve all purposes and that multiple quality models will be needed for different types of systems (MOODY, 2005). Furthermore, Hesse e Shneiderman (2007) states that a successful e-health system lies in merging the science of evidence-based medicine with the practice of user-centred research (HESSE; SHNEIDERMAN, 2007), yet most evaluation models do not explicitly focus on the end user.

1.2 PROBLEM

Since ICT-based healthcare can affect a patient's health status, there are a few quality guidelines for developing such systems. The ATA lists 13 guidelines for several kinds of systems from telerehabilitation services to videoconferencing-based telemental health (American Telemedicine Association, 2017).

Although some research approach the quality of ICT healthcare subject, most of them are related to a specific application (LEROUGE; GARFIELD; HEVNER, 2002; LEROUSE et al., 2004), because evaluating ICT-based healthcare systems is complicated due to of the complex nature of the healthcare domain (ANDARGOLI et al., 2017), and there is not a consensus on how to evaluate such systems, nor the better approach for attesting its quality (ANDARGOLI et al., 2017).

Related works are scarce, most software quality models tend to be generic, not focusing on a domain, and most models evaluate a portion of ISO/IEC25010 quality characteristics, failing to analyze software quality as a whole. Thus in practice, quality models are generic and there seems to lack of specific models customized for healthcare. Software quality evaluations of ICT-based healthcare systems in practice that are developed for the evaluation of just one specific system, not presenting a model that can be used to analyze similar systems in the same domain.

Thus this indicates a lack of one model that can evaluate software quality more aligned with ISO/IEC 25010 quality definition, i.e., a model that analyses a wider spectrum of quality characteristics. This also shows that there is a gap in evaluation models that are able to analyze other types of telemedicine and telehealth systems.

1.3 OBJECTIVES

In this work, we propose and validate a new model for evaluating software quality for telemedicine and telehealth systems, the AdEQUATE model (ALVES et al., 2015) from the viewpoint of end-users. This model is decomposed in quality characteristics systematically derived from ISO/IEC 25010 standard models and also in this research we develop a measuring instrument and an analysis guide for the evaluation.

1.3.1 Specific objectives

O1 Synthesize the state-of-the-art in software quality evaluation within the context of ICT-based healthcare systems;

O2 Develop a software quality evaluation model focused on the end user's perception;

O3 Develop a measuring instrument to collect data for the software quality evaluation;

O4 Apply the model to evaluate software quality of telemedicine and telehealth systems;

O5 Evaluate the model in terms of reliability and construct validity using the collected data in its applications [O4].

1.4 RESEARCH ALIGNMENT IN SOFTWARE ENGINEERING

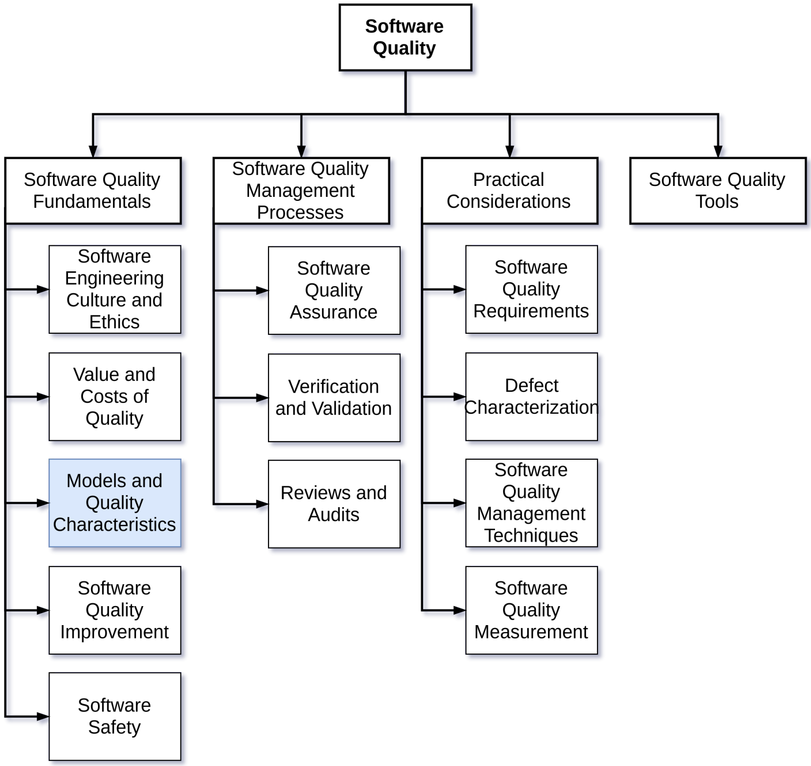
The specified objective of the software engineering research area of the PPGCC is: to train individuals capable of conducting the software development process and to investigate new methodologies, techniques, and tools for system design (PPGCC, 2017). Thus, in this context, the present research develops new knowledge relevant to the area of software engineering.

In the area of software engineering, software quality is a major topic (BOURQUE; FAIRLEY et al., 2014) as shown in Fig. 2. Within software quality, this work is defined inside the area of software quality fundamentals, in models and quality characteristics, because this research defines terminology for software quality characteristics and it differs from others software quality models in levels and number of characteristics and sub-characteristics (BOURQUE; FAIRLEY et al., 2014).

In this scope, we have created a software quality model that evaluates software quality from the end user's perspective based on the ISO/IEC 25010 international standard (WANGENHEIM et al., 2013).

The present work is applied in the healthcare domain encompassing systems that are within the contexts of telemedicine and telehealth,

Figure 2 – Breakdown of topics for the software quality knowledge area



Fonte: Adapted from (BOURQUE; FAIRLEY et al., 2014)

thus excluding m-health systems and other pervasive technologies that might be used in mobile devices.

1.5 METHODOLOGY

This research represents an applied research with an exploratory objective using quantitative and qualitative approach and bibliographic, documental and case study procedures as shown in Tab. 1. It is an applied research because it seeks to solve the stated problem (the gap in the current field of software quality evaluation models) (KOTHARI, 2004). It is also a qualitative research for it uses qualitative data gathered by the developed measuring instrument (KOTHARI, 2004). Likewise is an exploratory research that can hypothesize a reason for the quality degree of a telemedicine or telehealth software system (KOTHARI, 2004).

Table 1 – Research classification

Criteria	Classification	Justification
Nature	Applied	Regarding the general objective, that aims to develop a software quality evaluation model that could be applied to evaluate any telemedicine and telehealth software system (KOTHARI, 2004).
Objective	Exploratory	This research presents features of exploratory research, because it is related to the analysis of the state-of-the-art; surveying and case studies (YIN, 2013)
Approach	Quantitative & qualitative	Regarding the case studies that encompass data analysis of quantitative and qualitative data (YIN, 2013)
Procedures	Bibliographic, documental, case study	Regarding the methods applied to achieve our goals we use bibliographic procedures for objectives O1 and O2, and case studies in O3 and O4.

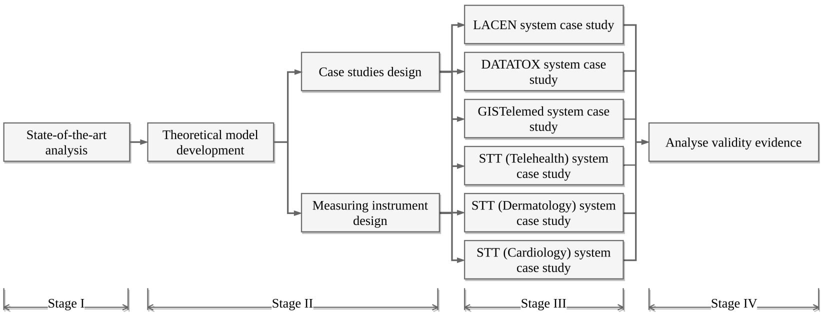
Fonte: Based on (KOTHARI, 2004)

To achieve the defined objectives, this research uses a multi-method approach, divided into 4 stages. The first stage regards the techniques and steps of the state-of-the-art synthesis, the second shows the methods used in the model development, the case study and measuring instrument design; The third stage regards the methods used in the execution of all case studies, and the final stage the methods used in the evaluation of the model, as illustrated in Fig. 3.

1.5.1 Stage I: State of the art analysis

The first stage aims to analyze the state-of-the-art in software quality evaluation models for telemedicine and telehealth systems (objective O1) encompassed a contextualization of software evaluations models focusing on international standards and the ICT-based healthcare domains focusing in telemedicine and telehealth. After the contextualization we have used Kitchenham (2004) method to discover the state-of-the-art in software quality evaluation models for telemedicine

Figure 3 – Research methodology



Fonte: Author's creation

and telehealth systems.

Thus in this stage, we defined 6 research questions related to which models exist to systematically evaluate software quality focusing on the end user's perception. We also defined inclusions and exclusions criteria to determine whether an article is analyzed or excluded from the state-of-the-art. A defined search string is executed in defined data sources and data is extracted and analyzed.

As result of this stage, we have synthesized the state-of-the-art of software quality evaluation models for telemedicine and telehealth systems and threats to the validity of this synthesis is discussed.

1.5.2 Stage II: Developing AdEQUATE

The second stage has the objective to develop both software quality model and its measuring instrument (Objectives O2 and O3). In this stage, after having selected consolidated state-of-the-art software quality evaluation standards and methods, we have systematically developed the AdEQUATE model and designed the evaluation case studies.

We have used the GQM approach (BASILI; CALDIERA; ROMBACH, 1994) for the theoretical model development, decomposing the model's goal, evaluate software quality of telemedicine and telehealth systems from the viewpoint of end-users, into questions, e.g., what is the effectiveness degree perceived by the system end-users. Those questions are based on the ISO/IEC 25010 standard models user perceivable characteristics (WANGENHEIM et al., 2013). To analyze each question, we propose 2 metrics and measures that facilitate the analysis of data collected by the measuring instrument.

Such instrument has been developed using Kasunic (2005) method. Thus we have defined the questionnaire items, selected the format of each questionnaire item and specified wording, designed the questionnaire sequence and overall layout and developed ancillary documents.

In the same phase we have designed each case study using Yin (2013) methodology, defining their goal as to evaluate software quality of 6 telemedicine and telehealth systems: The LACEN system, DATATOX system, cardiology, telehealth and dermatology modules of STT, GISTelemed; All using the AdEQUATE model's measuring instrument.

As result, the AdEQUATE model has been developed and the case studies defined, ready to be applied in Stage III.

1.5.3 Stage III: Case Study Execution and Analysis

This stage aims to collect data for the AdEQUATE model evaluation (Stage IV) and to evaluate the software quality of 6 different systems used in the Brazilian public healthcare.

The case studies defined in Stage II were executed using Yin (2013), using an online survey tool (SCHMITZ, 2017). For the cardiology, dermatology and telehealth modules of STT case studies we invited users via e-mail, and for the DATATOX, LACEN, and GISTelemed case studies we also have personally invited users because such systems have fewer active users.

In a third step after defining (Stage II) and executing the case studies, we have analyzed the data of each case, using AdEQUATE's metrics to assess the overall software quality of each system.

1.5.4 Stage IV: Evaluating AdEQUATE

The objective of this stage is to evaluate the reliability and construct validity of the AdEQUATE model. The data gathered in Stage III are used to evaluate reliability, a degree in which a set of items are measuring a single quality factor,i.e., the perception of software quality; and construct validity, its ability to measure what it proposes. Therefore, we adopted DeVellis (2016) method.

To evaluate reliability validity, we have used the Cronbach's α analysis (CRONBACH, 1951), and to evaluate construct validity we have used Spearman's correlation matrix and item-total correlation. As result of this stage, we have evaluated the AdEQUATE model's validity and discussed possible threats to the validity.

1.6 CONTRIBUTIONS

This work has produced scientific, technological, and social contributions.

1.6.1 Scientific contributions

This work has two major scientific contributions in the software engineering area. One contribution is the synthesis of the state-of-the-art in software quality evaluation models for telemedicine and telehealth systems (Chap. 3). This result also indicated a gap in the current field of software quality evaluation models, which this research aims to cover.

Another contribution is the AdEQUATE model itself, a software quality evaluation model developed for evaluation telemedicine and telehealth systems from the end-user point of view (Chap. 6).

1.6.2 Technological contributions

The main technological contribution of this work is the AdEQUATE model's measuring instrument – a 68 item questionnaire - that enables the analysis of any telemedicine and telehealth system from the viewpoint of the end-user in a useful way. Hence supporting the evaluation and development of telemedicine and telehealth systems through the evaluation of their quality.

During the application of the AdEQUATE model (Chap. 5), we also developed a dataset of end-user opinions of telemedicine and telehealth systems used in the Brazilian public health care context. This dataset is publicly available[1] enabling other researchers to further analyze the perception of end-users of the software quality of ICT-based healthcare systems.

[1] http://www-periodicos-capes-gov-br.ez46.periodicos.capes.gov.br/index.php\\?option=com_phome

1.6.3 Social contributions

The results of this research, whenever applied to any telemedicine or telehealth system, can assess the degree of the software quality as perceived by its end users. This can act as a compass pointing what should be improved in order to achieve sufficient software quality.

Contributing to the improvement of the software quality of telemedicine/telehealth system, the AdEQUATE model, can help to reduce costs, because by pointing out specific quality characteristics to be improved, it saves the cost a full system refactoring, and more important to reduce risks to the patient's health status, the fundamental principle of healthcare because it brings effectiveness and lowers the risk imposed to patient's health status.

1.7 WORK STRUCTURE

This work is divided into 7 chapters. In the following chapter, we describe the theoretical background to facilitate the understanding of the main concepts used in this research. In chapter 3 we provide an overview on the state-of-the-art of software quality evaluation models for telemedicine and telehealth systems. Chapter 4 presents the step-by-step development of the software quality model AdEQUATE.

In Chap. 5 we present the applications of the model, evaluating a total of 6 telemedicine and telehealth systems.

The validation of the model is presented in Chap. 6, and conclusions are given in Chap. 7

2 THEORETICAL BACKGROUND

In this chapter, we provide an overview of the concepts approached in this master thesis. In Sec. 2.1, we introduce the concept of software quality from the perspective the most used standard in the industry – ISO/IEC 25010 (LAMPASONA et al., 2012) and how to evaluate software quality using software quality models. Section 2.2 characterizes the ICT-based healthcare domain, its main stakeholders, and comments on its benefits.

2.1 SOFTWARE QUALITY

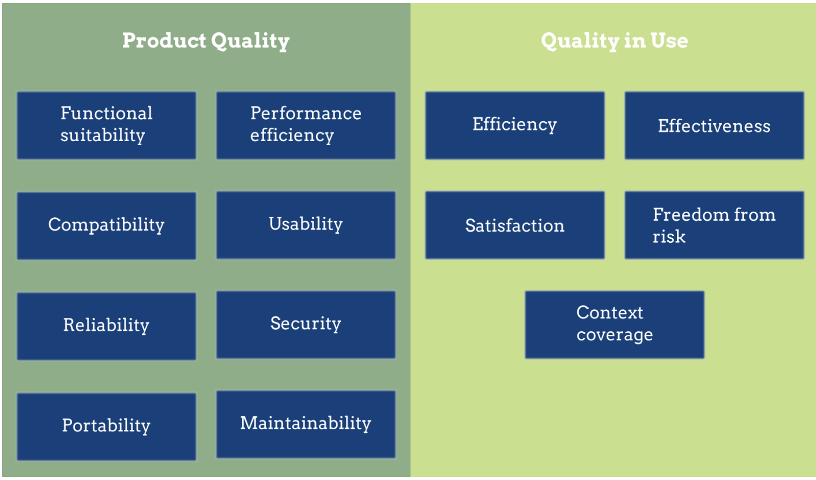
Software quality is defined as the degree to which a system satisfies the stated and implied needs of its various stakeholders, adding value to the final product (International Organization for Standardization, 2014). Due to the complexity of software systems, multiple factors impact on this degree, from how often the software system is updated (KHOMH et al., 2012), to how the code is reviewed (KEMERER; PAULK, 2009) among other factors. In order to further detail and specify the definition of software quality, several quality models have been proposed to support quality control and to measure how a software system is operating with quality (WAGNER, 2013). Among those models, the most prominent is the ISO/IEC 25010 standard (LAMPASONA et al., 2012).

2.1.1 ISO/IEC 25010 system and software quality models

The most used software quality models are defined in the ISO/IEC 25010 standard (LAMPASONA et al., 2012), shown in Fig. 4 (the full description of ISO/IEC 25010 is given in Annex A). The standard decomposes the abstract concept of software quality, into quality characteristics, which in some cases are further subdivided into sub characteristics (a sub-characteristic can sometimes be divided into sub-sub characteristics) (International Organization for Standardization, 2014).

ISO/IEC 25010 defines two software quality models: the product quality model (Fig. 4 left): that is decomposed into characteristics related to static software properties, and the quality in use model (Fig. 4 right), decomposed into characteristics related to outcomes of interaction with a system.

Figure 4 – ISO/IEC 25010 quality characteristics



Fonte: Adapted from (International Organization for Standardization, 2014)

The quality in use model consists of 5 quality characteristics: efficiency, effectiveness, satisfaction, freedom from risk, and context coverage, and 9 sub characteristics (whose definitions are given in Annex A).

Efficiency and *effectiveness* are the only quality characteristic that is not further decomposed into sub-characteristics. They are defined as the accuracy and completeness with which users achieve specified goals, and the resources expended in relation to effectiveness, respectively.

Satisfaction is defined as the degree to which user needs are satisfied when a product or system is used in a specified context of use.

Freedom from risk is a characteristic that is defined as the degree to which a product or system mitigates the potential risk to economic status, human life, health, or the environment.

Context coverage uses definitions of all other quality in use characteristics, and is defined as the degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in both specified contexts of use and in contexts beyond those initially explicitly identified.

The product quality model comprises 8 quality characteristics: functional suitability, reliability, performance efficiency, usability, security, compatibility, maintainability and portability, and 31 sub-characteristics (whose definitions are given in Annex A).

Functional suitability is the degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions, and *performance efficiency* is the performance relative to a number of resources used under such conditions. Additionally, the degree to which a system, product or component performs specified functions under specified conditions for a specified period of time is defined as *reliability*.

Compatibility is the degree to which a product, system or component can exchange information with other products, systems or components, and/or perform its required functions while sharing the same hardware or software environment.

Usability, in ISO/IEC 25010, uses definitions of characteristics from quality in use model and it is defined as the degree to which a product or system can be used by specified users to achieve specified goals with, *effectiveness*, *efficiency* and *satisfaction* in a specified context of use.

Security is defined as the degree to which a product or system protects information and data so that persons or other products or systems have the degree of data access appropriate to their types and levels of authorization.

Both *maintainability* and *portability* use definitions from quality in use model, and respectively defined as degree of effectiveness and efficiency with which a product or system can be modified by the intended maintainers; and the degree of effectiveness and efficiency with which a system, product or component can be transferred from one hardware, software or other operational or usage environment to another.

The product quality and quality in use models are useful for specifying requirements, establishing measures, and performing quality evaluations (International Organization for Standardization, 2014). The quality measures established are metrics associated with quality characteristics (or sub characteristics) (International Organization for Standardization, 2014) used to analyze the quality-related properties of a software system. However, to perform a quality evaluation,i.e., to determinate the quality degree of a software product, an evaluation model uses one or more measurement methods, defined as a logical sequence of operations used to quantify properties with respect to a specified scale (International Organization for Standardization, 2014), used to collect data in order to analyze the adherence to established quality measures (International Organization for Standardization, 2014; International Standard Organization, 2010;

LAMPASONA et al., 2012; BISCOGLIO; MARCHETTI, 2014).

2.1.2 Other software quality models

Other than the models presented in the ISO/IEC 25010 standard, software quality models have been used both in the academy and software industry.

One of the first software quality models (CÔTÉ; SURYN; GEORGIADOU, 2007) was the McCall model (MCCALL; RICHARDS; WALTERS, 1977), that aims to provide guidelines about how to specify how software quality is desired while specifying system requirements. McCall, Richards e Walters (1977) lists 11 characteristics: correctness, reliability, efficiency, integrity, usability, maintainability, testability, flexibility, portability, reusability, and interoperability.

A contemporary of McCall's model was the Boehm's model (BOEHM; BROWN; LIPOW, 1976). An interesting feature of Boehm's model is that differently from McCall it decomposes its quality characteristics hierarchically. First Boehm, Brown e Lipow (1976) decomposes quality into 3 characteristics: As-is utility, maintainability, and portability. As-is utility is decomposed into efficiency, reliability, and human engineering; Maintainability is decomposed into testability, understandability, and modifiability; portability is decomposed into device-independence and self-containedness.

The ISO/IEC 25010 standard itself is an update and revision of the former international standard ISO/IEC 9126 (International Organization for Standardization, 2014). The previous standard, ISO/IEC 9126, presented a product quality model that decomposed quality in 6 quality characteristics (ROCHA; MALDONADO; WEBER, 2001), 2 less than the updated standard's model.

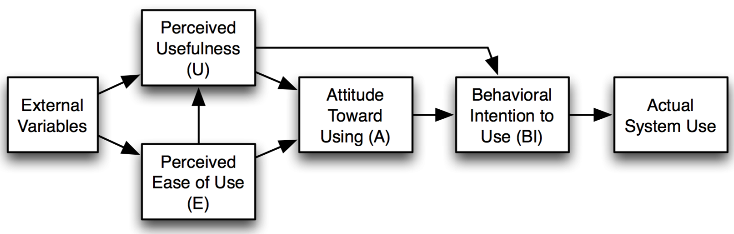
For the Dromey model (DROMEY, 1995), a model should explicitly link quality characteristics to the software code. In this approach structures related to the programming language or programming paradigm to software quality attributes, suggesting the use of ISO/IEC 9126.

In 1989 Davis, Bagozzi e Warshaw (1989) proposed the TAM, whose goal was to formulate better measures for predicting and explaining the use of technologies. This model was proven powerful and today is the most widely applied model to evaluate users' acceptance and usage of technology (VENKATESH, 2000). This model predicts and explains use of technology analyzing 2 variables: Perceived Usefulness

(U) – the degree to which a person believes that using a particular system would enhance his or her job performance (DAVIS, 1989) – and Perceived Ease of Use (E), that in contrast, is defined as the degree to which a person believes that using a particular system would be free of effort (DAVIS, 1989).

The TAM model is illustrated in Fig. 5. It holds that external variables (for instance a standard that imposes the use of the technology, or if the use of the technology will improve the user's curriculum) have an impact on U and E of the user. This vision of U and E generates an attitude toward using (A), and this A and U defines the behavioral intention to use (BI) that finally predicts and explains the actual system use.

Figure 5 – TAM model



Fonte: Adapted from (DAVIS; BAGOZZI; WARSHAW, 1989)

Other models have extended TAM and proposed other variables to the original TAM. For instance Venkatesh et al. (2003) uses four variables: Performance expectancy, effort expectancy, social influence, and facilitating conditions.

2.1.3 Software quality evaluation from the end-user viewpoint

Software quality evaluations are used to support decision-making (BOURQUE; FAIRLEY et al., 2014). With the increasing sophistication of software, questions of quality go beyond whether or not the software works to how well it achieves measurable quality goals (BOURQUE; FAIRLEY et al., 2014). Decisions supported by software quality measurement include determining levels of software quality (notably because models of software product quality include measures to determine the degree to which the software product achieves quality goals) (BOURQUE; FAIRLEY et al., 2014).

In this context, since the 1970's, it is believed that user involvement in system development ensures system success (HE; KING, 2008). Furthermore, the involvement of end-users in software quality aspect of the software, this including the evaluation of software quality, improves the quality of the system (BANO; ZOWGHI, 2015; FOSTER; FRANZ, 1998).

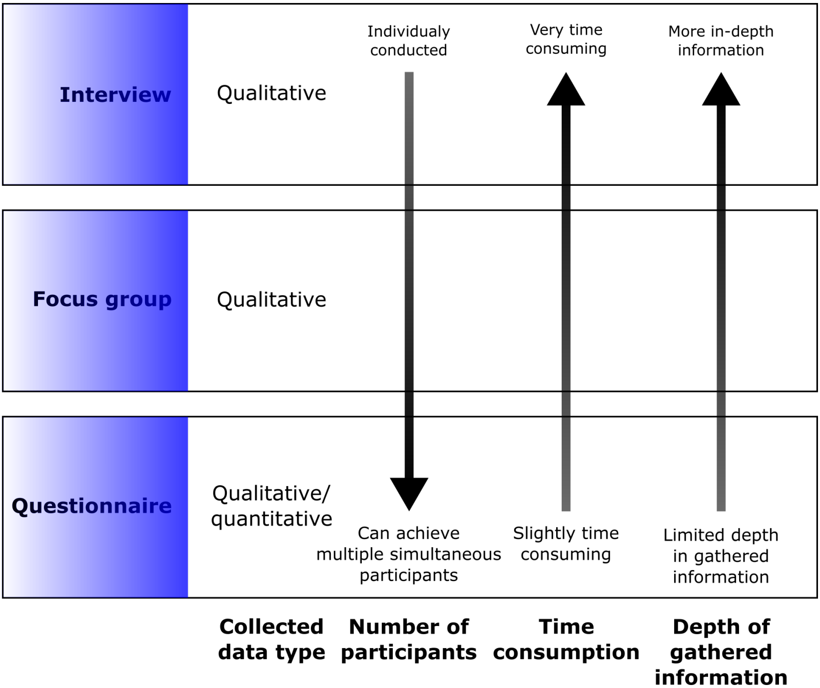
Common measurement methods when analyzing end user data are focus groups, interviews and questionnaires (ADAMS; COX, 2008) (Fig. 6), among others. Focus groups and interviews are based on a verbal communication and spoken narratives (RITCHIE et al., 2013), whilst focus groups are defined as an exercise in group dynamics (STEWART; SHAMDASANI, 2014; RITCHIE et al., 2013), interviews are conducted individually (RITCHIE et al., 2013). Interviews exist in a variety of forms ranging from formal interviews, through the Internet, over the telephone, or in face-to-face interaction, to more informal conversations conducted for research purposes, for example, as a part of ethnographic fieldwork (BRINKMANN, 2014).

Both interviews and focus groups generate qualitative data (COMPEAU; TASHAKKORI; TEDDLIE, 2003), that is more complex to analyze than quantitative data, generated by questionnaires (LONGHURST, 2003). However, generally, interviews (and focus groups in a lesser degree) provide more in-depth information in comparison to questionnaires (COMPEAU; TASHAKKORI; TEDDLIE, 2003). In turn, questionnaires are less time consuming and can easily achieve more participants than the other two approaches (DÖRNYEI; TAGUCHI, 2009; COMPEAU; TASHAKKORI; TEDDLIE, 2003; LONGHURST, 2003).

Questionnaires are easy to construct, extremely versatile, and uniquely in a form that is readily used (DÖRNYEI; TAGUCHI, 2009). In fact, any written instruments that present respondents with a series of questions or statements to which they are to react either by writing out their answers or a selecting from among existing answers could be considered a questionnaire (DÖRNYEI; TAGUCHI, 2009).

Regarding questionnaires, there is 2 formats of responses in a questionnaire (KASUNIC, 2005). The open-ended response format gathers qualitative data where respondents create their own answers to the question in their own words. There are no answer choices provided from which they select their responses. On the other hand, the close-ended response format gathers quantitative data in two different ways: Ordered response options and unordered response options (KA-

Figure 6 – Measurement instruments comparison



Fonte: Based on (LONGHURST, 2003; COMPEAU; TASHAKKORI; TEDDLIE, 2003)

SUNIC, 2005). In close-ended ordered with ordered response options, respondents choose between provided response options represented in a graduated scale (KASUNIC, 2005). Differently, with unordered response options, respondents choose from among discrete, unordered response options (KASUNIC, 2005).

In the software quality evaluation context, there exists different (standardized) measuring instruments to measure specific quality factors, such as e.g., the SUS BROOKE et al., for the subjective assessments of usability (BROOKE et al., 1996) from the end-user point of view. Another approach similar to SUS, is the CSUQ, that measure users' satisfaction with the usability of computer systems (LEWIS, 1995).

An essential issue with respect to the measuring instruments is to assess if an evaluation model and its measuring instrument truly measure what they pretend to measure (construct validity) (DEVEL-

LIS, 2016) and if the same measuring process provides the same results (reliability) (CARMINES; ZELLER, 1979). To analyze reliability and construct validity, DeVellis (2016) advocates three statistical methods: Cronbach's α, Spearman correlation and item-total correlation analysis. To assess reliability, the Cronbach's α approach, defined by Cronbach (1951), proposes an α coefficient that shows the degree in which a set of items are measuring a single factor.

Regarding construct validity, Spearman correlation is based on a matrix of correlation coefficients that indicates the degree of correlation between a pair of factors (COHEN, 2013); Item-total correlation analysis indicates the correlation of a factor with all other factors.

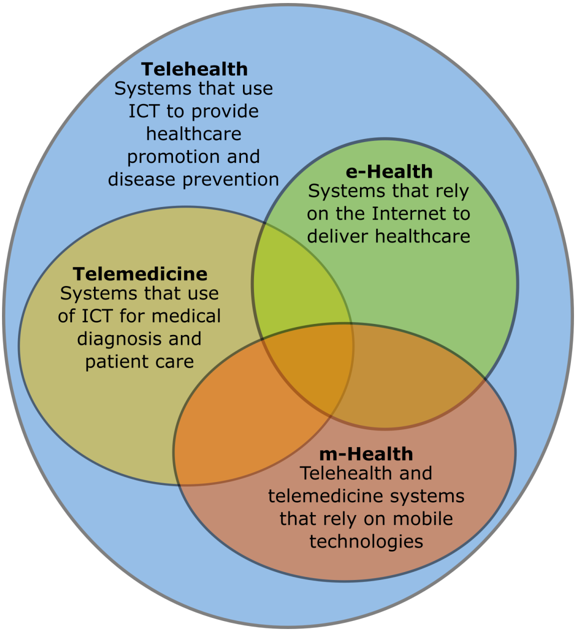
2.2 ICT-BASED HEALTHCARE

The evolution of ICT played a major role in the development of all ICT health domains (BASHSHUR et al., 2011). Beginning with basic telecommunication (origins of telemedicine), followed by expanding the scope of ICT based healthcare (telehealth), networking (e-health), and most recently, the "personalization" of ICT-based healthcare (m-health). These domains, "telemedicine", "telehealth", "e-health" and others, are not well defined, often overlap, and have created a confusion that hinders the conceptual development of ICT-based healthcare (BASHSHUR et al., 2011; American Telemedicine Association, 2013; EYSENBACH, 2001). Nonetheless, the goal of ICT-based healthcare, regardless of ICT domain, is to improve quality by increasing adherence to guidelines, enhancing disease surveillance, and decreasing medication errors (CHAUDHRY et al., 2006).

In this context, ICT health domains, shown in Fig. 7, could be further refined. Telemedicine, that has more than 100 different definitions (SOOD et al., 2007), is commonly defined the use of telecommunications for medical diagnosis and patient care (SCANNELL; PEREDNIA; KISSMAN, 1995; BASHSHUR et al., 2011; BASHSHUR; REARDON; SHANNON, 2000; SMITH; GRAY, 2009). It involves the use of ICT as a medium for the provision of medical services to sites that are distant from the provider. The concept encompasses everything from the use of standard telephone services through high speed, wide bandwidth transmission of digitalized signals in conjunction with computers, fiber optics, satellites and other sophisticated peripheral equipment and software (SCANNELL; PEREDNIA; KISSMAN, 1995).

Conceptually, telemedicine is to telehealth is what medicine is to

Figure 7 – Definition of overlaping ICT-based health domains



Fonte: Based on (BASHSHUR et al., 2011; SCANNELL; PEREDNIA; KISSMAN, 1995; BASHSHUR; REARDON; SHANNON, 2000; SMITH; GRAY, 2009; SHOWELL; NOHR, 2012; American Telemedicine Association, 2013; OH et al., 2005)

health (BASHSHUR et al., 2011). Hence, telehealth is considered to have a broader scope than telemedicine and is defined as systems that use ICT to provide health care promotion and disease prevention (KOCH, 2006; EYSENBACH, 2001).

Although there is not a general consensus about the meaning of e-health (SHOWELL; NOHR, 2012; American Telemedicine Association, 2013), it may refer to health applications that rely on electronic processing and the Internet (BASHSHUR et al., 2011) or an even more inclusive reference to any computer usage in healthcare (OH et al., 2005).

The term m-health was introduced into the literature in 2003 (ISTEPANIAN; LACAL, 2003) in response to the vast expansion of mobile communication technology and its perceived usefulness in facilitating access to healthcare (BASHSHUR et al., 2011). Thus, m-health incorporates the preceding domains of telemedicine, telehealth, and e-health as long as their applications depend on mobile communications and network technologies (BASHSHUR et al., 2011).

2.2.1 Telemedicine & telehealth

Encompassing various domains, ICT-based healthcare systems may vary on its components, functionalities, applications and technologies (BASHSHUR et al., 2011). As noted in section 1.4, this work focuses on the telemedicine and telehealth domains, therefore e-health and m-health domains, are beyond this research scope, and the remainder of this section focuses on telemedicine and telehealth applications.

Telemedicine systems have 3 different aspects called dimensions: functionality, applications and technological configurations (BASHSHUR et al., 2011), as shown in Tab. 2, that are composed on a number of components.

Table 2 – Dimensions and componentes of telemedicine systems

Telemedicine dimension	**Dimension components**
Functionality	Consultation
	Diagnosis
	Mentoring
	Monitoring
Applications	Specialty
	Disease
	Site
	Treatment
Technology	Synchronicity
	Network
	Connectivity

Fonte: Based on Bashshur et al. (2011)

The first dimension, functionality, incorporates all aspects of the medical care process, including activities in prevention, diagnosis, treatment, follow-up, and rehabilitation (BASHSHUR et al., 2011). The application dimension includes processes of care across virtually all basic medical specialties, as well as sub specialization based on disease entities, sites of care, and treatment modalities. The last dimension describes the architecture in which the system is based upon.

The *consultation* functionality is implemented in teleconsultation systems or modules. It is defined by a consultation occurring between two (or more) healthcare providers (such as physicians) as well as between provider and patient (BASHSHUR; REARDON; SHANNON, 2000; BASHSHUR et al., 2011; SOOD et al., 2007).

The *diagnosis* functionality is implemented in telediagnosis systems or modules. It is defined by when a specialist physician (such as, e.g., a radiologist, a pathologist, or a cardiologist), relying on transferred images, records, and/or laboratory results diagnoses a patient (BASHSHUR et al., 2011; SOOD et al., 2007). Moreover, consultation and diagnosis are not mutually exclusive functionalities, hence both can occur in the same moment (BASHSHUR et al., 2011).

The *monitoring* functionality is implemented in telemonitoring systems or modules. It is defined by the monitoring of patients that are home-bound because are chronically ill, recently discharged or otherwise requires continuous observation (such as, e.g., patients with congestive heart failure, chronic obstructive pulmonary disease, asthma, and/or diabetes) (BASHSHUR, 2001).

The *mentoring* functionality is implemented in telementoring systems or modules. It is defined by the remote guidance typically by surgeons and other specialists to other surgeons performing new or complex procedures (BASHSHUR et al., 2011).

Regarding the application dimension, the *specialty* component is related to which medical specialty the system provides. The trend over the years has been one of increasing specialization and subspecialization in medicine in response to scientific and technological advances. This trend is reflected in telemedicine, in which basic specialties include content areas such as teledermatology, teleradiology, and telepsychiatry (BASHSHUR et al., 2011). Some systems are further specific and are related to a particular *disease* component. In this context, we can observe systems for specific diseases such as e.g., diabetes, stroke, and post-traumatic stress disorder.

There are diverse medical specialties in telemedicine applications, and to list all specialties and diseases is beyond the scope of this research. However, Tab. 3 and 4 show the description of specialties and diseases that, have been considered as effective to the traditional face-to-face medical practice.

Other components of the telemedicine application dimension are *site* and *treatment*. Telemedicine systems can be used in different sites of care, including the intensive care unit, outpatient psychiatry, the emergency department, and the home (BASHSHUR et al., 2011). Concerning treatment, Some programs have been organized around speci-

Table 3 – Medical specialties used in telemedicine

Medical specialty	Purpose of telemedicine application
Surgery	Neurosurgical consultation
	Fine-tissue laparoscopic manipulation
	Intercontinental telelaparoscopic adrenalectomy
	Training and mentoring to community surgeons
	Nephrectomies in porcine models
Ophthalmology	Screening for cytomegalovirus retinitis
	Macular disorder referral
	Diabetic retinopathy-screening examinations
	Diagnosis of treatment-warranted retinopathy of prematurity
Pathology and microbiology	Second-opinion service on frozen sections
	Digitalization of frozen section and tissue-smear preparations of brain tumors
	Consultation over difficult cases
	Diagnosis of infectious diseases during missions
Oncology	Teleoncology visits
Gastroenterology	Telerehabilitation system for dysphagia
Dermatology	Teleconsultation in remote countries
	Virtual lesion clinic assessment
Cardiology	Telemanagement services for patients with chronic heart failure and chronic obstructive pulmonary disease
	Telediagnosis for cardiac surgery
	Remote fetal echocardiograms
	Home TV channel for teleguiding chronic heart failure patients
	Neonatal cases with suspected cardiopathy surveillance
	Remote sonography in a pediatric cardiology center
Dentistry	Interceptive orthodontic services provided to disadvantaged children in order to reduce malocclusion severity
	Consultation between US Armed Forces dentists and specialists on patients' status
Psychiatry and neurology	Intervention delivery by geographical information system
	Post-traumatic stress disorder treatment by videoconferencing
	Depression-related disorders
	Teleconferencing to diagnose and assess potential psychiatric disorders in asylum seekers, refugees, and migrants
	Referral diagnosis replay
	Therapy for obsessive–compulsive disorders

Continues on the next page

fic treatment modalities such as rehabilitation (e.g., speech/language

Table 4 – Medical specialties used in telemedicine

Medical specialty	Purpose of telemedicine application
Genetics	Screening for potential genetic syndromes using telemedicine
	Video conferencing to deliver genetic counseling services
	Genetic counseling for hereditary breast and/or ovarian cancer
Geriatrics	A remote-site management
	Mobility-impaired-patient management
	In-home telerehabilitation of elderly
Allergy and immunology	Promote immunology healthcare and allergy education
Anesthesiology	Use of Teleconsultations to facilitate anesthesia procedures
Clinical laboratory sciences	ICT utilization to execute tests on biological specimens health evaluation, diagnosis, prevention, or treatment of illnesses and deficiencies
Emergency medicine	Teleconsultation applications on emergency sites
Endocrinology	Teleconsultation applications to provide endocrinology healthcare
Family medicine	Teleconsultation applications to provide family practice
Forensic medicine	Use of ICT to exchange pathological data for virtual forensic applications
Toxicology	Use of ICT in the investigation, diagnosis and management of suspected poisoning cases

Fonte: Based on Cerbo et al. (2015), Thomas (2012), ONA - Organização Nacional de Acreditação (2006), Ostojic et al. (2005), Grundy et al. (1977), Pavlopoulos et al. (1998), Marcin et al. (2005), Norris et al. (2002), Thali et al. (2003), Sood et al. (2007)

pathology, physical therapy) and pharmacy (BASHSHUR et al., 2011).

The network dimension is decomposed into 3 components: synchronicity, network, and connectivity. Regarding *synchronicity*, telemedicine can be either synchronous (i.e., real time) referring to the concurrent presence of interacting participants located at different places; or asynchronous (store-and-forward), in which the participants do not interact in real time (BASHSHUR et al., 2011).

Network design/configuration includes three modalities: Virtual Private Networks, the open Internet, and social networks, in which information is posted and shared. The three modalities substantially vary in terms of security arrangements and the ability to protect confidential information (BASHSHUR et al., 2011).

Regardless of network and synchronicity, both may be enabled by two types of connectivity, wired and wireless, both of which now provide different levels of bandwidth and the attendant speed and resolution or quality of service (BASHSHUR et al., 2011).

Telehealth systems, as shown in Fig. 7, are more inclusive, encompassing other specialized concepts (LEROUSE et al., 2004). Therefore it is more generically defined than telemedicine. The 4 telehealth activities described in Tab. 5.

Table 5 – Telehealth activities

Activity	**Definition**	**Synchronicity**
Teleconsulting	Question and answer registered to solve doubts related to management, conduct and clinical procedures, healthcare actions, and questions related to the work process, based on scientific evidence but locally adequate.	Asynchronous and synchronous
Second formative opinion	Structured answers to questions originated from teleconsulting and selected using relevance and inclusion criteria from SUS directives. These answers are given based on bibliography and best clinical and scientific evidence	Asynchronous
Telediagnose	Diagnose support where exams are performed in one location and transferred to another location using ICT. The medical report will be given by a specialist physician	Asynchronous
Tele-education	Educative activities ministered from a location and transmitted via ICT.	Asynchronous and synchronous

Fonte: Based on Ministerio da Saúde (2015)

The *teleconsulting* activity can be performed either synchronously by electronic chat-rooms, web-conferencing or video conferencing; or asynchronously, using store-and-forward questions that have to be answered in 72 hours.

Having such a variety of possible applications, from transcontinental telesurgeries (JANETSCHEK; BARTSCH; KAVOUSSI, 1998) to answering questions about clinical cases (DAMASCENO et al., 2014), telemedicine and telehealth systems support a number of different stakeholders (LEROUGE; GARFIELD; COLLINS, 2012), such as represented in table 6.

Direct users can be further defined into 5 groups, as shown in Tab. 7. Different systems may specify different user classes within each group, relating users with their medical specialty, or sectors within the health institution.

Table 6 – Stakeholders of ICT based healthcare systems

Stakeholder Type	Description	Domain
Direct users	Physicians, nurses and other clinicians responsible for the direct care and to provide orientations to patients.	Telemedicine, and e-Health
Coordinators	Responsible for the selection and installation of equipment, overseeing scheduling, and general operations.	All
Administrators	Responsible for the establishment of policies and procedures.	All
Technical experts	Responsible for the installation configuration, and maintenance of equipment.	All
Researchers	Engaged in data collection and analysis for multiple projects.	All
Patients	User of medical services	Telemedicine, telehealth, and m-Health

Fonte: Based on LeRouge, Garfield e Collins (2012), Koch (2006)

Table 7 – Telemedicine and telehealth direct users

Stakeholder	Description	Example (Context)
Health-care provider	Analyses health data (e.g. medical images) to deliver healthcare	Dermatologist (teledermatology), cardiology (telecardiology), biochemist (clinical systems), toxicologist (poisoning control)
Health-care requesters	Requires the analysis of health data	Nurse, physician
Technical colaborators	Facilitates the sending of health related data and its screening and observation	Attendants, nurses
Education provider	Provides healthcare information about its area of expertise	Specialist physician
Education requesters	Requires information about healthcare	Nurse, physician

Fonte: Based on (LEROUGE; HEVNER; COLLINS, 2007; KOCH, 2006; SOOD et al., 2007; EKELAND; BOWES; FLOTTORP, 2010)

3 STATE-OF-THE-ART

This chapter presents the synthesis of the state-of-the-art on software quality models for the evaluation of telemedicine and telehealth systems from the end user's perspective. As defined in our methodology (Sec. 1.5), we performed a systematic literature review, whose steps are shown in the following sections.

3.1 DEFINITION OF THE REVIEW

The objective of this systematic review is to *synthesize the state-of-the-art in literature pertaining software quality evaluation models for ICT-based healthcare systems that focus on the end user's perception.* To achieve this objective, we analyze the following research questions:

RQ1 Which models exist to systematically evaluate software quality focusing on the end user's perception?

RQ2 Which quality and/or sub-quality characteristics are evaluated?

RQ3 How the data collection and analysis is operationalized?

RQ4 How these models have been developed?

RQ5 How these models have been validated?

RQ6 Which ICT-based healthcare domain the model focuses on?

3.1.1 Inclusion/Exclusion Criteria

In this research, we have focused, and defined the inclusion criteria, on articles that presented a model to evaluate software quality from the end user's perception.

We focused only on peer-reviewed articles, written in English, available via digital libraries published in the last 10 years (between January 2005 and December 2015).

We excluded any article not related to a software quality model, articles that do not present empirical study/evaluation, and articles that do not focus on the end user's perspective. Likewise, we have excluded any article presented a model customized for any of ICT-based health domain.

3.1.2 Data sources

Data sources have been chosen considering their coverage in the computing and medicine domain and their ability to handle advanced queries. The chosen repositories are ACM Digital Library, IEEE Xplore, Springer Link, Science Direct, PubMed and Wiley Online Library. All repositories are freely available through CAPES portal[1].

3.1.3 Quality criteria

In addition to our inclusion/exclusion criteria, we also superficially assessed the quality of the reported studies, considering only articles that provide substantial information on the evaluation model.

3.1.4 Search string

The search string is structured according to population, intervention and comparison (KITCHENHAM, 2004). We omitted, however, the outcome and context facet from the search string structure as our research questions do not call for a restriction of the results to a particular outcome or context. The keywords are shown in Tab. 8. The core concepts, representing population, intervention, and comparison, are derived from our research questions. We identified synonymous, related/broader/wider concepts and alternative spelling for each core concept. We did not include specific keywords from existing technologies or methods to capture the end user's perception of software quality to the population set of terms as this could have biased the search (KITCHENHAM, 2004).

The search string was calibrated to the specific syntax of every data source, without adding or removing information from the search string. The resulting search string for each data source is shown in Tab. 9.

[1] `http://www-periodicos-capes-gov-br.ez46.periodicos.capes.gov.br/index.php\\?option=com\_phome`

Table 8 – Search string keywords

Concept	Term
Software quality	“software quality”; “system quality”
Evaluation	evaluation; assessment; survey; appraisal; decision; assurance; model; method;,framework; scale;
end user	user; “end user”; “end user”;
ICT Health	e-health; telemedicine; telehealth; m-health

Fonte: Author’s creation

Table 9 – Specific search string per repository

Database	Search string
IEEE Xplore Digital Library	("software-quality"OR "system-quality") AND ("evaluation" OR "assessment"OR "survey"OR "model"OR "appraisal" OR "method") AND ("user"OR "end-user"OR "final-user") AND ("telemedicine"OR "telehealth"OR "e-health"OR "m-health") &pys=2005&pye=2015
Science Direct Digital Library	("software-quality"OR "system-quality") AND ("evaluation" OR "assessment"OR "survey"OR "model"OR "appraisal" OR "method") AND ("user"OR "end-user"OR "final-user") AND ("telemedicine"OR "telehealth"OR "e-health"OR "m-health") &date=2005-2015
ACM Digital Library	("software-quality"OR "system-quality") AND ("evaluation" OR "assessment"OR "survey"OR "model"OR "appraisal" OR "method") AND ("user"OR "end-user"OR "final-user") AND ("telemedicine"OR "telehealth"OR "e-health"OR "m-health") &dte=2005&bfr=2015
Springer Link Digital Library	("software-quality"OR "system-quality") AND ("evaluation" OR "assessment"OR "survey"OR "model"OR "appraisal" OR "method") AND ("user"OR "end-user"OR "final-user") AND ("telemedicine"OR "telehealth"OR "e-health"OR "m-health") &facet-start-year=2005&facet-end-year=2015
Wiley Online	("software-quality"OR "system-quality") AND ("evaluation" OR "assessment"OR "survey"OR "model"OR "appraisal" OR "method") AND ("user"OR "end-user"OR "final-user") AND ("telemedicine"OR "telehealth"OR "e-health"OR "m-health") &start-year=2005&end-year=2015
PubMed	("software-quality"OR "system-quality") AND ("evaluation" OR "assessment"OR "survey"OR "model"OR "appraisal" OR "method") AND ("user"OR "end-user"OR "final-user") AND ("telemedicine"OR "telehealth"OR "e-health"OR "m-health") &sy=2007&ey=2015

Fonte: Author’s creation

3.2 EXECUTION

The SLR was conducted in 2016 by the author, a Computer Science master's student. In the initial search, we found a total of 36,140 articles, summarized per data source in Tab. 10. From ACM Digital Library, Springer Link and Science Direct we selected only the 1,000 most relevant results, observing a lack of relevance afterward. From IEEExplore, Scopus and PubMed all returned articles were analyzed. As result, a total of 3110 articles were analyzed during the first stage.

Table 10 – Search results

Database	**Initial Result**	**1st Selection**	**2nd Selection**	**Relevant Studies**	
IEEE Xplorer	90	90	2	2	2
ACM Digital Library	4938	1000	12	6	5
Springer Link	2968	1000	16	8	3
Science Direct	28124	1000	44	19	12
Scopus	16	16	0	0	0
PubMed	4	4	0	0	0
Sum	36140	3110	74	37	22

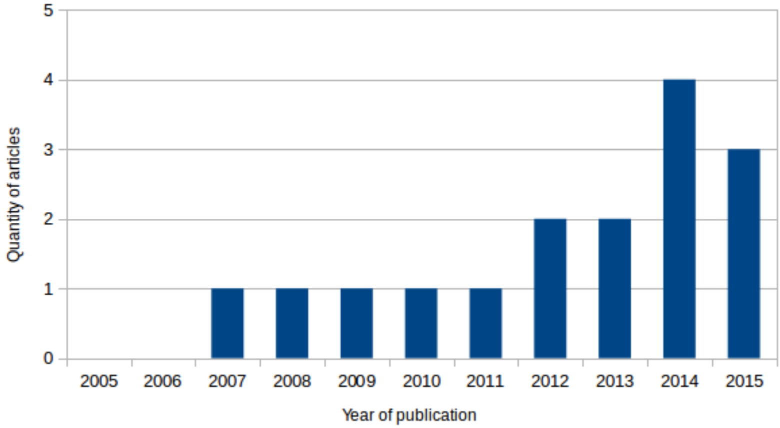
Fonte: Author's creation

During the first stage, the search results were quickly analyzed based on their title and short summary. The abstract was read only in case the title or summary did not provide enough evidence of any inclusion/exclusion criteria. Irrelevant and duplicate articles were removed. This stage left us with 74 potentially relevant articles. Most articles were eliminated because they did not present the quality characteristics which they analyzed or the quality evaluation was made from a perspective other than the end user's point of view.

During the second stage, we analyzed the articles' abstract and quickly looked for information on the quality characteristics evaluated. As result, 16 articles were identified as relevant articles. The other articles were removed, because they did not present sufficient information on how data was gathered and/or the quality characteristics evaluated.

The relevant articles are summarized in Tab. 11. In this context, the chart in Fig. 8 shows that the majority of relevant articles have been published since 2012. This shows that the interest in approaches to systematically evaluate software quality of ICT-based healthcare systems has been growing in the last four years.

Figure 8 – Distribution of studies per year of publication



Fonte: Author's creation

3.2.1 Data extraction

In accordance with the research questions, we systematically extracted information (shown in Tab. 12) in a spreadsheet from each article selected for analysis.

All articles deemed relevant were read thoroughly and data was extracted by the author. Data extraction was hindered in several cases by the way in which the studies were reported. Most articles lack sufficient detail about the theoretical basis, development methodology and how end users of the medical context were involved. In these cases, coherent data was extracted as much as possible.

3.3 DATA ANALYSIS

We have identified 16 articles presenting 15 different approaches to evaluate software quality from an end-user point of view. The following subsections 3.3.1, 3.3.2, 3.3.3, 3.3.4, 3.3.5, and 3.3.6 address each of the research questions.

Table 11 – Relevant articles

Title	Citation
From expert-derived user needs to user-perceived ease of use and usefulness: A two-phase mixed-methods evaluation framework	(BOLAND et al., 2014)
An Evaluation Framework for EU Research and Development e-Health Projects' Systems	(MAVRIDIS; KATRIOU; KOUMPIS, 2008)
Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology	(BROWN et al., 2013)
Development, implementation and evaluation of an information model for archetype based user responsive medical data visualization	(KOPANITSA; VESELI; YAMPOLSKY, 2015)
A usability and accessibility design and evaluation framework for ICT services	(SUBASI; LEITNER; TSCHELIGI, 2009
Adaptation of an Evaluation System for e-Health Environments	(SÁNCHEZ-PI; MOLINA, 2010)
An end user evaluation of query formulation and results review tools in three medical meta-search engines	(LEROY et al., 2007)
Ergonomic evaluation and design of a mobile application for maternal and infant health for smartphone users among lower-income class Filipinos	(NOCUM et al., 2015)
Evaluation of a Social Web-based Telehealth System	(DHILLON; WÜNSCHE; LUTTEROTH, 2012a)
Evaluation of a Web-Based Telehealth System: A Preliminary Investigation with Seniors in New Zealand Modeling the longitudinally of user acceptance of technology with an evidence-adaptive clinical decision support system	(DHILLON; WÜNSCHE; LUTTEROTH, 2012b)
Modeling the longitudinally of user acceptance of technology with an evidence-adaptive clinical decision support system	(JOHNSON; ZHENG; PADMAN, 2014)
KopAL Appointment user-interface - An Evaluation with Elderdly	(FUDICKAR; FAERBER; SCHNOR, 2011)
Evaluation of a guidelines-based e-health decision support system for primary health care in South Africa	(HORNER; COLEMAN, 2014)
Assessing the User Experience Design as a Persuasive Methodology in a Real World Sports Application	(PILLONI et al., 2013)
Usability evaluation of pharmacogenomics clinical decision support aids and clinical knowledge resources in a computerized provider order entry system: A mixed methods approach	(DEVINE et al., 2014)
User-oriented evaluation of a medical image retrieval system for radiologists	(MARKONIS et al., 2015)

Fonte: Author's creation

Table 12 – Specification of extracted data

Research question	Data item	Description
RQ#1	Reference	Reference of the article
	Name	Acronym or name of the model
RQ#2	Quality (sub) Characteristics	Quality (sub-) characteristic(s) that are evaluated
RQ#3	Data collection instrument	Type of instrument(s) used for data collection
RQ#4	Development methodology	Methodology used in the model development
	Evaliation method	Method(s) used to validate the model
	Data points	Size of the dataset used in the model's validation
RQ#6	Health ICT domain	ICT-based healthcare domain the model is focused on

Fonte: Author's creation

3.3.1 RQ1 – Which models exist to systematically evaluate software quality focusing on the end user's perception

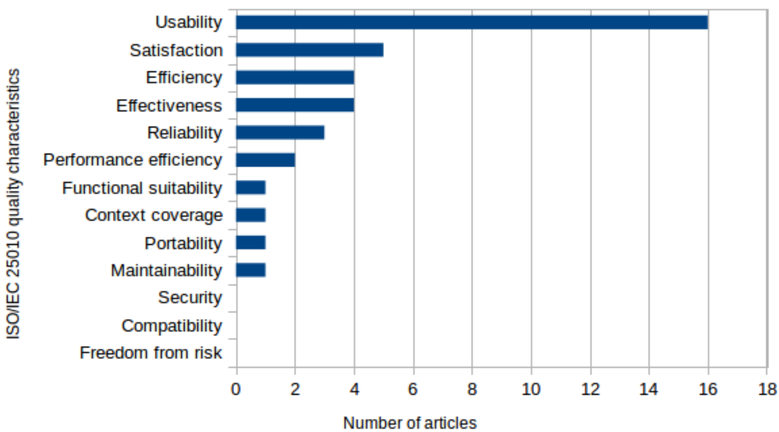
Most of the 15 models presented in the selected articles do not present a complete description that includes a name. Only Brown et al. (2013) names its model, the Health IT Usability Evaluation Model (Health-ITUEM). The Health-ITUEM was developed as an integrated model of multiple theories as a comprehensive usability evaluation model. The Health-ITUEM was originally developed by assessing the usability of an ICT system for scheduling nursing staff and adapted to the context of m-health.

The BOLAND et al. model combines an evaluation from the end user's perspective with the system's perspective (software quality requirements). This enables the selection of a suitable comparison system followed by a cognitive walk-through involving a task analysis and a comparison of interface design (BOLAND et al., 2014).

3.3.2 RQ2 – Which quality and/or sub-quality characteristics are evaluated?

Considering the selected 15 models, most of them do not define explicitly software quality, nor their decomposition of software quality into characteristics and consequently in sub-charactersitcs. The chart in Fig. 9 shows the number of ISO/IEC 25010 quality characteristics considered in the evaluation of the relevant articles. Hence, this chart shows that all articles have evaluated usability. The second most evaluated quality characteristic is satisfaction. The security, compatibility, and freedom from risk characteristics are not evaluated by any of the selected models. Regarding this descriptive statistics, we can see that the mode of evaluated characteristics is 1, hence, we conclude that the majority of software quality models evaluate just 1 software quality characteristic.

Figure 9 – Quantity of articles that evaluate ISO/25010 quality characteristics

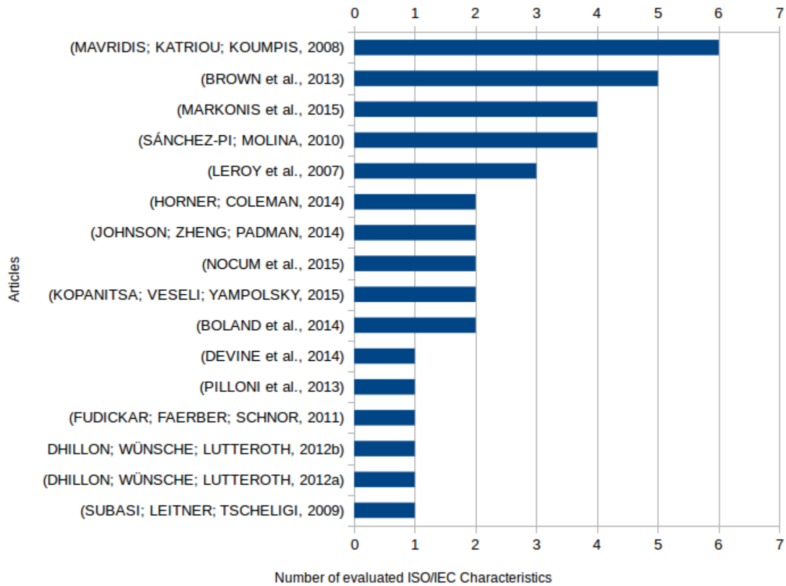


Fonte: Author's creation

A different perspective to address this research questions in to analyze how complete (or specific) the relevant articles approach ISO/IEC 25010 quality characteristics. The model described by (MAVRIDIS; KATRIOU; KOUMPIS, 2008) is the model that evaluates the highest number of different quality characteristics, however, it still does not evaluate all ISO/IEC 25010 characteristics (Sec. 2.1.1, Fig. 4).

Brown et al. (2013), in second place, evaluates 5, followed by Markonis et al. (2015) and Sánchez-Pi e Molina (2010) evaluating, both, 4. As expected due the analysis of the chart in Fig. 9, again the majority of models evaluates just one characteristic, and these models are described by Devine et al. (2014), Pilloni et al. (2013), Fudickar, Faerber e Schnor (2011), Dhillon, Wünsche e Lutteroth (2012a, 2012b), and Subasi, Leitner e Tscheligi (2009).

Figure 10 – Quantity of ISO 25010 quality characteristics used in selected articles



Fonte: Author's creation

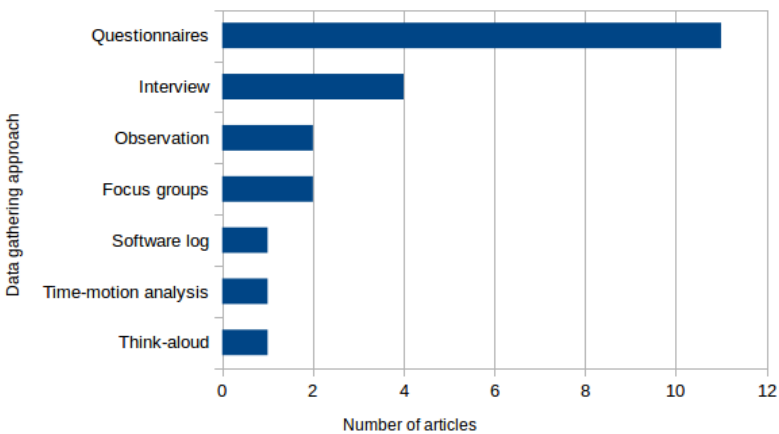
3.3.3 RQ3 – How the data collection and analysis is operationalized?

The processes of software quality evaluation described in the selected models are quite straightforward, and we can divide the evaluation processes into 2 categories: Models that preconise a software quality definition and decomposition before the execution evaluation itself, and models that are ready to use. The modes described by Devine et al. (2014), Pilloni et al. (2013), Horner e Coleman (2014), Fudickar, Faerber e Schnor (2011), Johnson, Zheng e Padman (2014), Agnisarman et al. (2017), Dhillon, Wünsche e Lutteroth (2012a, 2012b), Nocum et al. (2015), Leroy et al. (2007), Sánchez-Pi e Molina (2010), Brown et al. (2013) propose the use of a measuring instrument that is ready to be used as a part of a evaluation research methodology, such as, e.g., a case study.

The other category of models preconise that previous to the evaluation, the software quality characteristics should be defined in the context of the evaluation. Kopanitsa, Veseli e Yampolsky (2015), Boland et al. (2014), Mavridis, Katriou e Koumpis (2008) propose such models. An exception to these categories is proposed by Markonis et al. (2015). It is similar to models that do not predefine software quality characteristics, but differently, it defines an evaluation stage before the execution of the measuring instrument.

Regarding the data gathering approach, the chart in Fig. 11 shows that questionnaires are the most popular approach, followed by interviews, used in 11 and 4 models respectively. Software log, time-motion analysis, and think-aloud are used in just one model.

Figure 11 – Data gathering approaches used by number of articles

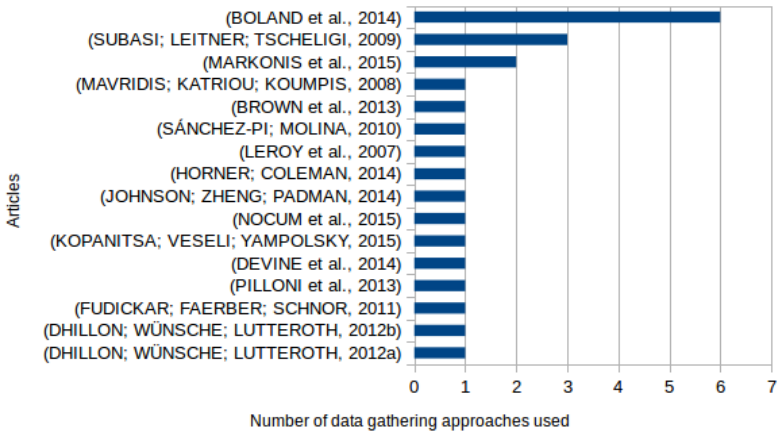


Fonte: Author's creation

The perspective is shown in the chart in Fig. 11 do not show however that, some models use multiple data gathering approaches.

This view on this research question is presented by the chart in Fig. 12. Thus we can see that Boland et al. (2014) uses the most different number of approaches, but this is a rather uncommon behavior because 13 authors use just one approach.

Figure 12 – Number of data gathering approaches used in the models



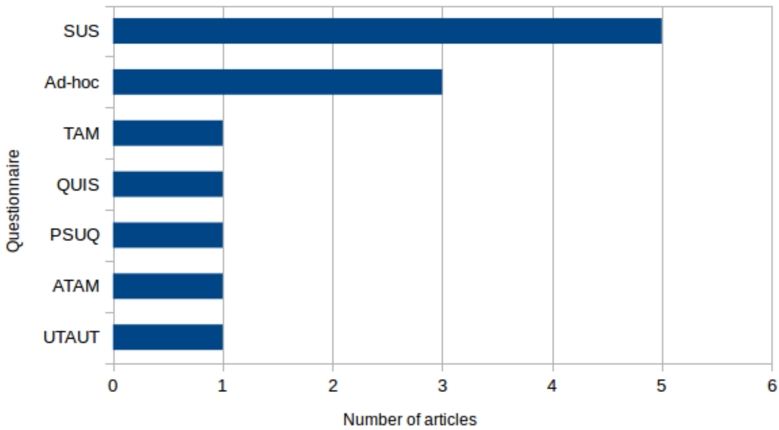
Fonte: Author's creation

Since almost all relevant articles used at least one questionnaire, we have specified each questionnaire in the chart shown in Fig. 12. The chart in Fig. 13 shows that the SUS – a simple, ten-item scale giving a global view of subjective assessments of usability (BROOKE et al., 1996) – has been used by the majority of the relevant articles, not only as a post-test questionnaire but also as a part of models and frameworks. Three studies, (SÁNCHEZ-PI; MOLINA, 2010; LEROY et al., 2007; NOCUM et al., 2015) propose and use ad-hoc questionnaires that have not been systematically developed, nor specified the methodology used in its development.

The UTAUT questionnaire is also used in two articles, (BOLAND et al., 2014) that uses it as a post-test questionnaire complementary to SUS, but (CIMPERMAN; BRENČIČ; TRKMAN, 2016) extends it to better capture the acceptance and use of technology of home telehealth systems.

Mavridis, Katriou e Koumpis (2008) use the ATAM questionnaire as a part of their framework to enlist the quality characteristics to be evaluated, and it is not to be confused with using the TAM questi-

Figure 13 – Number of questionnaires used in the models



Fonte: Author's creation

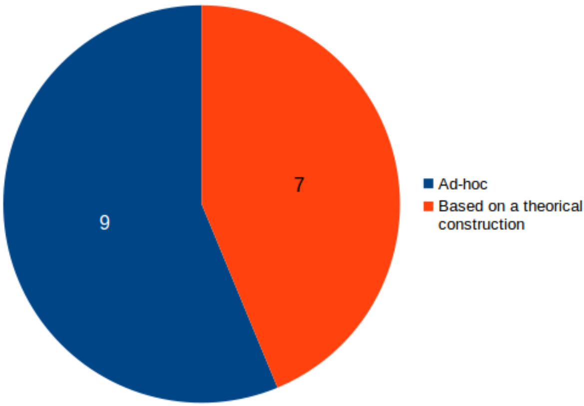
onnaire, as is used by (JOHNSON; ZHENG; PADMAN, 2014), as a post-test questionnaire that evaluates perceived usefulness and perceived ease of use, which are hypothesised to be fundamental determinants of user acceptance (DAVIS, 1989). A similar model proposed by (HADJI; DEGOULET, 2016) is the ITPAM2, that uses TAM, UTAUT, and others to evaluate quality in each context or deployment phase of a clinical information system.

Another approach similar to SUS is the Computer System Usability Questionnaire (CSUQ), a variant of PSUQ that measure users' satisfaction with the usability of computer systems (LEWIS, 1995), applied by (AGNISARMAN et al., 2017) as a post-test questionnaire. The original PSUQ, used in (DEVINE et al., 2014), has a similar goal and it also applied as a post-test questionnaire. In a similar manner, Markonis et al. (2015) uses the QUIS as a complement of SUS.

3.3.4 RQ4 – How these models have been developed?

We identified that the articles were very unclear about the methodology used in the development stage of the described models. Thus we do not consider that any reported a systematic methodology used to develop the model. In general, the models seem to be developed in an ad-hoc manner. This behavior is illustrated by the pie chart in Fig. 14. Regardless we have considered that seven models did, however, specified a software quality model that supports the definitions used in the articles, presented in the pie chart as the seven articles based on a theoretical construction.

Figure 14 – Proportion of models ad-hoc developed and those based on a theoretical construction



Fonte: Author's creation

3.3.5 RQ5 – How these models have been validated?

In order to analyze if and how the models have been validated, we observed the methods and techniques used to analyze the validity of the models. We identified that the most of the models do not explicitly define any validation criteria, therefore their validity is very limited. The articles that mention a validation – Brown et al. (2013), Leroy et al. (2007), Johnson, Zheng e Padman (2014) – propose the model and are partially validated, as Brown et al. (2013) analyzes reliability, and Leroy et al. (2007), Johnson, Zheng e Padman (2014) analyze construct validity.

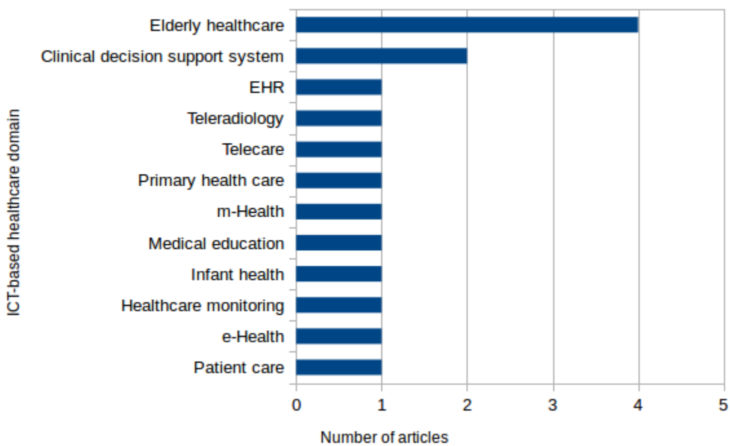
The validity of the model has been analyzed through some case or pilot studies, applying the model to evaluate a pre-existence ICT based healthcare system. For data analysis for validation of the research, Brown et al. (2013) uses inter-rater reliability measure, Leroy et al. (2007) uses variance analysis (ANOVA) and Johnson, Zheng e Padman

(2014) uses correlation and regression.

3.3.6 RQ6 – Which ICT-based healthcare domain the model focuses on?

Most models have evaluated software systems not necessarily in a comprehensive way, but focusing on specific kinds of ICT-based healthcare systems, application domains or user categories. The majority of the relevant articles focused their evaluations on elderly healthcare, followed by clinical decision support system as shown in the chart in Fig. 15. However, we can notice that there is no common practice or a tendency as several other ICT-based healthcare domains are evaluated.

Figure 15 – ICT health domains in selected articles



Fonte: Author's creation

3.3.7 Discussion

Regarding software quality models, ISO/IEC 25010 stands out as a reference cited in many articles analyzed in this systematic literature review. As a part of the Systems and software Quality Requirements and Evaluation (SQuaRE) (International Organization for Standardization, 2014) group of international standards, the software quality characteristics of ISO/IEC 25010 could be analyzed using metrics defined in ISO/IEC 25022 and ISO/IEC 25023. Such as, e.g. rate of user error to analyze effectiveness, time to complete a task to analyze efficiency. However, this SRL shows that those approaches are not used in the current state of the art, specially regarding the end-user's perspective.

Analyzing the data gathered in this systematic literature review, we have observed in the first selection over 3000 articles, and selected only 16 relevant studies. Thus we have concluded that there is a lack of software quality evaluation models for ICT-based healthcare systems that focus on the end-user evaluation. This characterizes a gap in the current field of software quality evaluation, therefore, a gap in the field of software engineering. Furthermore, the models are not systematically developed, causing difficulties to other researchers analyze the viability of the use of such models.

Considering the quality characteristics defined in the ISO/IEC 25010 standard models, usability is seldom evaluated. This shows that in the domain of ICT-based healthcare systems usability is of great importance for end-users. On the other hand, the absence of models evaluating security, compatibility, and freedom from risk shows that the perception of end-users does not facilitate the analysis of these characteristics. Another gap detected by this review is that no model evaluates all software quality characteristics defined by the standard, showing that the end-user perception has limitations.

The models used basically three methods to gather data from the end-user perception, interviews, focus groups and questionnaires, and often these approaches are combined. However, questionnaires are the most used tool, because in the way they have been developed it enables researchers to gather quantitative data and to analyze a larger number of end-users. However, these questionnaires are developed in an ad-hoc manner not in a systematic fashion, again difficulties further analysis by the scientific community. Nevertheless, the greatest gap in the models presented is the lack of validation evidence. Further work in the context of software evaluation models for ICT-based healthcare systems focused on the end-users perception should be validated at least in terms of reliability.

This review has shown an increased interest in the elderly healthcare field. This might be a concern with the aging population of the world, but as various other healthcare domains were tackled by relevant articles, we do not conclude that a particular domain is more evaluated than others.

3.3.7.1 Threats of validity

Inherent to every SRL is the bias relative to the scientific research publication itself. The majority of results are related to successes rather than failures (UNTERKALMSTEINER et al., 2012). This threat might have excluded from this review, models that have week validation evidence. Yet, we believe that this threat was mitigated by including a research question related to the validity of the models.

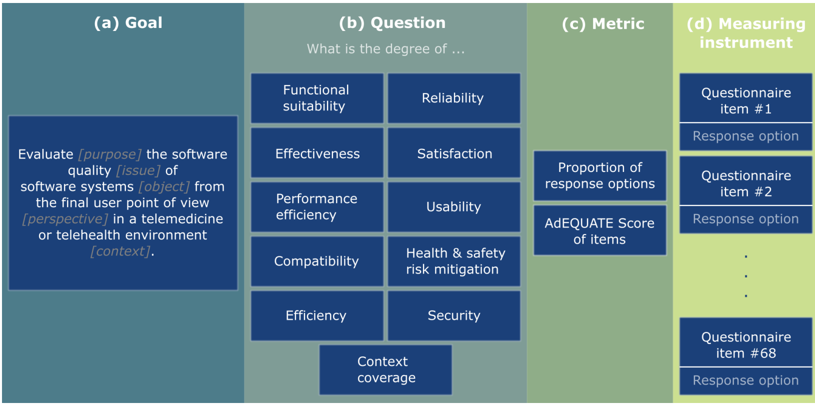
In the execution stage, there is a threat that relevant articles would not be encountered (UNTERKALMSTEINER et al., 2012). We have mitigated this threat by including synonyms to every search terms. Another mitigating strategy was to consider multiple data sources.

Another threat is the personal opinions of the researcher (UNTERKALMSTEINER et al., 2012). These opinions might influence the article selection. We have mitigated this threat by defining specific exclusion criteria, in which the selection was based.

4 THE ADEQUATE MODEL

In this chapter, we present the results of the development of the AdEQUATE model and its measuring instrument. Following the GQM method (BASILI; CALDIERA; ROMBACH, 1994), we defined a goal in terms of purpose, issue, object, perspective and context (KOZIOLEK, 2008) as shown in Fig. 16.

Figure 16 – Decomposition of AdEQUATE



Fonte: Author's creation

We have systematically decomposed the goal (Fig. 16.a) into analysis questions related to quality characteristics as shown in Fig. 16.b. Hence, the goal is decomposed into 11 analysis questions.

We choose to evaluate software quality based on the ISO/IEC 25010:2011 standard as it defines quality models considered the most used in software industry (LAMPASONA et al., 2012). We have also used its predecessor (ISO/IEC 9126:2001), and TAM as well, because it is used by related works as identified in the literature review (KOPANITSA; VESELI; YAMPOLSKY, 2015; NOCUM et al., 2015).

Therefore the AdEQUATE characteristics (Fig. 16.b) are a subset of ISO/IEC 25010 standard models characteristics, excluding the portability and maintainability characteristics because AdEQUATE models consider only characteristics that could be perceived by end-users and are considered relevant for telemedicine and telehealth systems (WANGENHEIM et al., 2013).

Since the AdEQUATE model is focused on, besides the perspective of end-users, the evaluation of systems within the telemedicine and telehealth domain, we have used the relevant articles in the state-of-the-art synthesized in Chap 3 to define each quality characteristic.

- Effectiveness: Degree of accuracy within appropriate precision degree and completeness with which users achieve specified goals (International Organization for Standardization, 2014; Clinical Laboratory Standards Institute, 2008; SOOD et al., 2007; LEROUGE; GARFIELD; COLLINS, 2012).

- Efficiency: Resources expended in relation to effectiveness.

- Satisfaction: Degree to which user needs are satisfied when a product or system is used in a specified context of use.

- Health & safety risk mitigation: Degree to which a product or system mitigates potential the direct of indirect risk to people health in the intended contexts of use (LEROUGE; GARFIELD, 2013; MEA, 2001).

- Context coverage: Degree to which a product or system can be used with effectiveness, efficiency, health, and safety risk mitigation and satisfaction in both specified contexts of use and in contexts beyond those initially explicitly identified.

- Functional suitability: Degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions.

- Performance efficiency: Performance relative to a number of resources used under such conditions.

- Reliability: Degree to which a system, product or component performs specified functions under specified conditions for a specified period of time.

- Compatibility: Degree to which a product, system or component can exchange information with other products, systems or components, and/or perform its required functions while sharing the same hardware or software environment.

- Usability: Degree to which a product or system can be used by specified users to achieve specified goals with, *effectiveness*:, *efficiency*: and *satisfaction*: in a specified context of use.

- Security: Degree to which a product or system protects information and data so that persons or other products or systems have the degree of data access appropriate to their types and levels of authorization.

Thus, based on the definition presented above, we decompose our goal *evaluate the software quality of software systems from the final user point of view in a telemedicine or telehealth environment* into 11 analysis questions (AQ):

AQ1: What is the degree of *functional suitability*;

AQ2: What is the degree of *reliability*;

AQ3: What is the degree of *effectiveness*;

AQ4: What is the degree of *satisfaction*;

AQ5: What is the degree of *performance efficiency*;

AQ6: What is the degree of *usability*;

AQ7: What is the degree of *compatibility*;

AQ8: What is the degree of *health and safety risk mitigation*;

AQ9: What is the degree of *efficiency*;

AQ10: What is the degree of *security*;

AQ11: What is the degree of *context coverage*.

To measure each analysis question, we have defined 2 metrics (Fig. 16.c): One is the proportion of end-user perception by quality degree. For instance, considering the analysis question "what is the degree of functional suitability", this metric would analyze the proportion of end-users that consider the as high degree of functional suitability, and the proportion of end-users who do not.

The other metric is a score from 0 to 100 that consider the end-user opinion, similar to the SUS approach to quantification of usability (BROOKE et al., 1996; BROOKE, 2013). For instance, considering the analysis question "what is the degree of functional suitability", this metric would analyze the end-user opinion to produce a general view of the functional suitability of the system.

To operationalize the analysis of end-user perception, the AdEQUATE model differs from ISO/IEC 25010 and proposes a standardized measuring instrument in a form of questionnaire that considers specificities of telemedicine and telehealth domains.

4.1 THE ADEQUATE QUESTIONNAIRE

Since the objective is focused on the end users' perspective, we have characterized the target audience as direct users of telemedicine and telehealth systems (Tab. 6)

To assess the degree of AdEQUATE characteristics we defined questionnaire items as beliefs questions – questions regarding the opinion of the respondent – that were systematically created according to every quality characteristic of the AdEQUATE model.

The questionnaire items were developed using adaptations from the ISO/IEC 25010 subcharacteristics themselves and based on literature using TAM (DAVIS; BAGOZZI; WARSHAW, 1989), the SUS (BROOKE et al., 1996) and other related works (KUO; LIU; MA, 2013; EGEA; GONZÁLEZ, 2011; PARASURAMAN, 2000; LIU; MA, 2006; DUCEY, 2013; CHAU; HU, 2002).

Regarding the structure of the questionnaire items, all of the responses are closed ended, thus the AdEQUATE questionnaire provides quantitative data of end-user perspective. We chose response options written in a positive manner to facilitate the interpretation of the questionnaire item and to mitigate errors (SAURO; LEWIS, 2011). The response options for every questionnaire item are:

- Totally agree
- Agree
- Totally disagree
- Disagree
- Not applicable: The participant does not interact with this aspect of the system and considers this item not applicable to the specific system being evaluated. For example, if the item is related to interoperability but the respondent does not need another system to operate it should select this option.
- Do not understand the item: The participant did not understand the item, (e.g., because of the wording) and thus prefers not to answer.
- Do not know: The participant understood the item, but does not have the knowledge or experience to respond it.

We have decided to use a 4-point Likert scale because there is no evidence that a mid-point improved reliability or validity of the scale (GARLAND, 1991), and there is evidence that a 4-point scale helps users to express their opinion (CHANG, 1994).

The options "not applicable", "do not understand the item", and "do not know" enables end-user to precisely answer the items, and enables a finer evaluation of the questionnaire (SAURO; LEWIS, 2011).

4.1.1 Questionnaire organization

The questionnaire is organized as follows:

1. Title
2. Introduction: Explains the objective of the research, ethical issues and the benefits of the study
3. Guidelines: General guidelines to answer the questionnaire
4. Questionnaire items

We decided to use an electronic format using the LimeSurvey tool (SCHMITZ, 2017), and present all items in a single page using radio buttons for every item response option. The AdEQUATE questionnaire is available online[1].

4.1.2 Analysis model

We have designed the AdEQUATE questionnaire to be used in any telemedicine or telehealth system. However, even so, these ICT-based healthcare domains may require further customization. Thus every questionnaire item is combined with a help text. The help text facilitates the comprehension of the questionnaire items (SAURO; LEWIS, 2011), and by customizing the help text to the specific telemedicine/telehealth context, the help text brings a perspective to which consider the questionnaire item (SAURO, 2015).

Additionally, we propose an analysis model to facilitate data analysis. Mathematically, we have defined the vector $\vec{D}_{1,11}$, and each position within this vector represents one of the 11 quality characteristics of the AdEQUATE model.

[1] `https://arquivos.ufsc.br/d/ce70c8e83e/`

Let $O = \{$"*Totally agree*", "*Agree*", "*Disagree*", "*Totally disagree*", "*Don't know*", "*Don't understand*", "*Not applicable*"$\}$ be the set of all response options. And let $\vec{R}_{1,N}$ be a vector of all responses as $R_i = \{o|o \in O\}$ and $0 < i \leq N$ being N the number of responses.

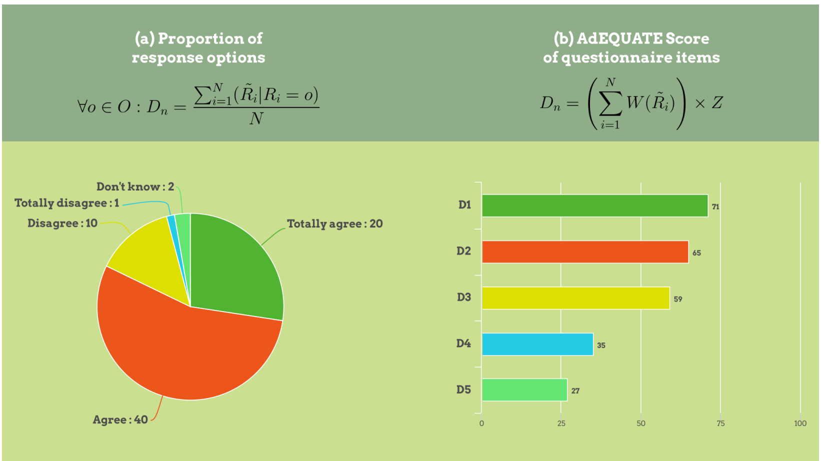
Furthermore we have defined the weighting function W (Eq. 1) and the normalizing value Z (Eq. 2).

$$W(o) = \begin{Bmatrix} 3, & if\ o = \text{“}Totally\ agree\text{”} \\ 2, & if\ o = \text{“}Agree\text{”} \\ 1, & if\ o = \text{“}Disagree\text{”} \\ 0, & otherwise \end{Bmatrix} \tag{1}$$

$$Z = \frac{100}{N \times W(\text{“}Totally\ agree\text{”})} = \frac{100}{N \times 3} \tag{2}$$

In this context, the equations are shown in Fig. 17 can analyze both AdEQUATE metrics.

Figure 17 – AdEQUATE metrics



Fonte: Author's creation

Thus, for every D_n, we have the proportion of every response option o in relation to N (17.a), and the scoring system represented in Fig. 17.b. Additionally Fig. 17 shows 2 illustrative hypothetical charts. The pie chart on the left represents a hypothetical view using descriptive statistics of a quality characteristic as per the proportion of response options metric of the AdEQUATE model. In the other hand, the chart on the right shows the view of the AdEQUATE score for five quality characteristics for a hypothetical evaluation.

A spreadsheet that operationalizes these formulae and enable the analysis of data gathered from an AdEQUATE questionnaire is available at[2].

To use the AdEQUATE model to evaluate a telemedicine/telehealth system, a researcher has to consider 2 steps:

1. Analyze the system direct users
2. Customize help texts for the system domain and direct users.

Researchers have to analyze the main stakeholders of the system in order to customize the help text that is combined with every questionnaire item. Having instantiated the LACEN questionnaire, the researcher is ready to execute the evaluation and, later, analyze the data using the proposed analysis model using the evaluation spreadsheet.

[2] `https://arquivos.ufsc.br/d/ce70c8e83e/`

5 APPLICATION OF THE ADEQUATE IN THE EVALUATION OF TELEMEDICINE AND TELEHEALTH SYSTEMS

This chapter shows applications of the AdEQUATE model in pilot case studies. The goal of these case studies are:

- To analyze the software quality degree of these systems;
- To collect data to analyze the reliability and construct validity of the AdEQUATE model.

We have conducted a series of case studies evaluating 6 telemedicine and telehealth systems in the context of the STT and ABRACIT as shown in Tab. 13.

Table 13 – AdEQUATE pilot tests

System (Module)	Activities	Direct users
LACEN	LIS	Health-care providers, Health-care requesters, Technical colaborators
DATATOX	PCS	Health-care providers, Health-care requesters, Technical colaborators
STT (Dermatology)	Teledermatology	Health-care providers, Health-care requesters, Technical colaborators
STT (ECG)	Telecardiology	Health-care providers, Health-care requesters, Technical colaborators
STT (Telehealth)	Teleconsulting	Education requesters, Education providers Technical colaborators
STT (GISTelemed)	Telemedicine GIS	Health-care providers, Health-care requesters, Technical colaborators

Fonte: Author's creation

The Santa Catarina State Integrated Telemedicine and Telehealth System is a set of information systems, services, and modules that provide access to healthcare and healthcare education in the Santa Catarina state, Brazil (BECKHAUSER et al., 2016). Such systems include a telemedicine and a telehealth modules, both available as web applications and as applications for mobile devices (WALLAUER et al., 2008; NOBRE; WANGENHEIM, 2012), and a Laboratory Information System (LIS) used by the Central Laboratory of Public Health (LACEN) (ALVES et al., 2014).

In use since 2005, the STT is available in all municipalities of the Santa Catarina state, being accessed by hundreds of users including medical doctors, nurses, community health agents, and biochemists. By the end of 2015, STT/SC has managed over four and a half million examinations of various specialties (INACIO et al., 2014).

The ABRACIT is an organization that represents toxicology information centers in Brazil, of which 9 use the DATATOX system to manage poisoning cases.

The following sections present each of the conducted case studies conducted by the author, the GISTelemed evaluation was performed within the context of this research but by other master degree student in Inacio (2016). The discussion and threats of the validity of all studies are presented in the end of this chapter.

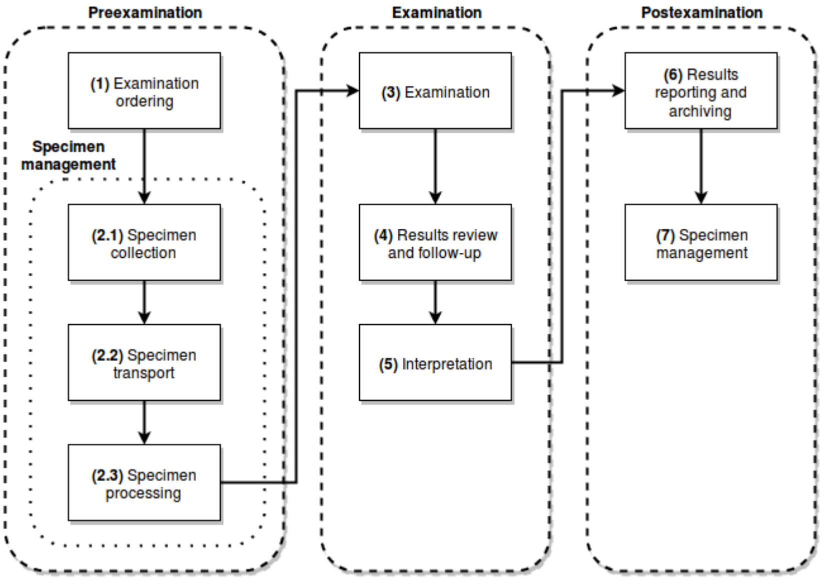
5.1 LACEN CASE STUDY

The first case study analyses the LACEN system (ALVES et al., 2016b), a responsible for the clinical analysis of high complexity laboratory examinations such as Dengue fever or AIDS. The LACEN system provides features for executing tests on biological specimens collected from patients, providing information for a health evaluation, diagnosis, prevention, or treatment of illnesses and deficiencies (ONA - Organização Nacional de Acreditação, 2006). Data gathered from patients, tests, results, and quality control procedures are managed by the LACEN system that organizes the laboratory workflow according to three main phases (Preexamination, Examination, and Post examination) and its subdivisions (Clinical Laboratory Standards Institute, 2008) as shown in Fig. 18.

The LACEN system is used by different direct users:

- biochemists as health-care providers
- physicians as health-care requesters
- laboratory technicians as technical collaborators

Figure 18 – Workflow of the LACEN System



Fonte: Author's creation

The LACEN system implements five core functionalities:

- Test ordering: Manage information regarding specimen orders, including provider (e.g. identification, contact information), patient (e.g. identification, previously known conditions, demographics), and order (e.g. requested tests, urgency status, authorizations). This functionality covers the process (1) of the *Preexamination* phase.

- Specimen accessioning and processing: Optimize specimen collection by providing instructions and guidelines for the responsible staff, attributing unique identifiers to specimens, and linking specimens to institutional operations. This functionality implements the specimen management process (2.1, 2.2, and 2.3) of the *Preexamination* phase.

- Analytic phase managing: Mitigate errors in the examination phase by managing workloads, showing the standardized operating procedures, and associating mandatory components necessary for testing with individual test records. This functionality implements the examination process (3) of the *Examination* phase.

- Result entry and validation: Guide analysts into providing or correcting high-quality test results by automated and/or manual entry following established security and certification levels, enabling biochemist supervisors to check and validate content prior to its liberation. This functionality implements the result review process (4) and interpretation (5) of the *Examination* phase.
- Result reporting: Provide the exam requester with a medical report containing the obtained results, following a standardized template specifying units of measurement, references, confidence intervals, and clinical methodologies associated with the performed tests. This functionality covers the results reporting process (6) in the *Post examination* phase.

The specimen management process (Fig. 18.7) does not need to have an associated functionality implemented since it encompasses particular laboratory policies and practices as to discard specimens or to store specimens for later review (Clinical Laboratory Standards Institute, 2008; SEPULVEDA; YOUNG, 2013).

The figure 19 shows a screenshot of the LACEN system. We can see an interface that enables the specimen accessioning and processing.

5.1.1 Study definition

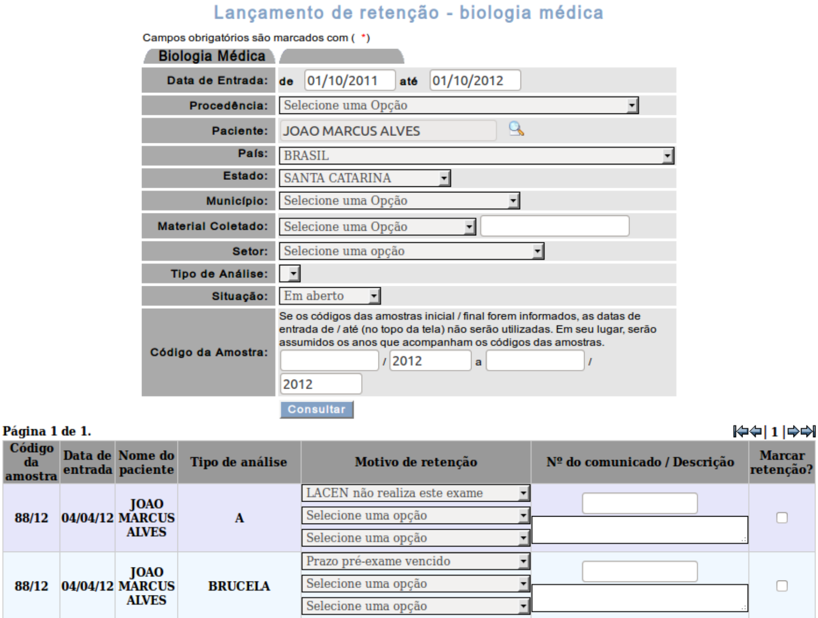
The objective of this case study is to evaluate the quality of the LACEN system in its context of use – the public Unified Health System in the State of Santa Catarina, Brazil from the end-user perspective. Therefore we adopted the AdEQUATE model, and in order to instantiate the model for the specific system, the AdEQUATE model's questionnaire has been revised and customized with counterexamples specific to the LACEN laboratory routine. The questionnaire applied in this case study is available online at [1].

5.1.2 Study execution

During the execution phase, we applied the customized AdEQUATE questionnaire in 2015, collecting data from end users of the

[1] https://arquivos.ufsc.br/d/ce70c8e83e/

Figure 19 – Screenshot of the LACEN System



Fonte: Author's creation

LACEN system who have been using it in their daily working routine. Users located at the LACEN's main facility have been personally invited; users from other facilities were invited through the login screen available on the system. From a total of 352 invitations, we obtained 69 valid responses, a response ratio of 19.60%.

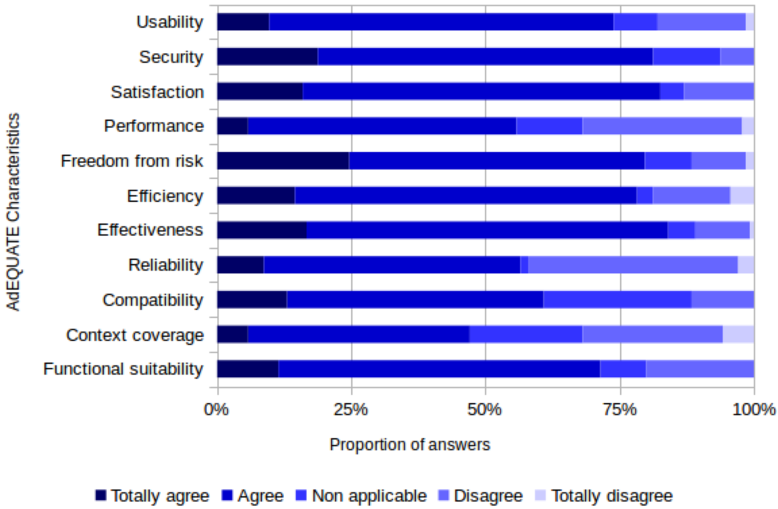
The demographics of respondents is as follows: 60 direct users (51 technical collaborators, and 9 health-care providers), and9 administrators. As patients do not access the LACEN system directly, they were not included.

The majority of the users that participated in the study have a considerable experience: 47 respondents use the system for over 3 years, 18 respondents use the system for between 1 and 3 years, and only 4 respondents use the system for less than 1 year.

5.1.3 Data analysis

Questionnaire items were grouped by its characteristics and using AdEQUATE's metrics we have analyzed the data using descriptive statistical techniques, as shown in Fig. 20 and 21.

Figure 20 – AdEQUATE proportion of opinion response option – LACEN responses



Fonte: Author's creation

The chart in Fig. 20 shows the percentage of each response option for each characteristic according to AdEQUATE first metric, the proportion of each response option (Fig. 17.a).

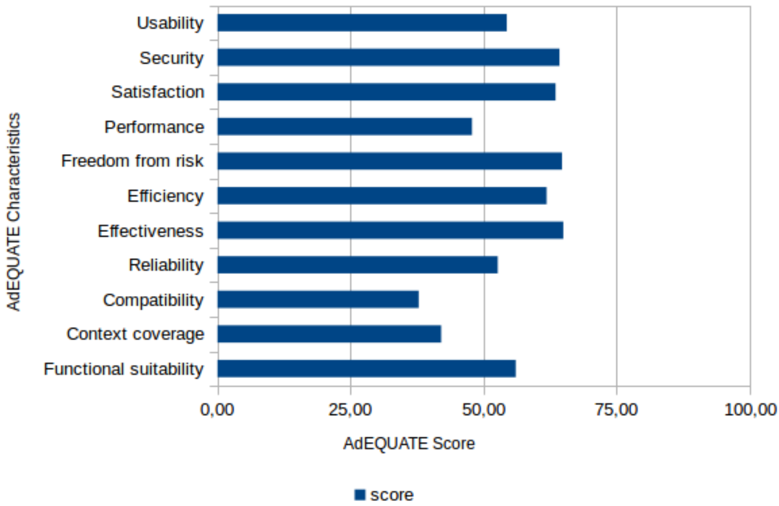
The majority of characteristics have been evaluated positively, with more than 50% being either "totally agree" or "agree". Only 1 characteristic had a percentage inferior to 50%: "Context coverage".

The characteristics evaluated with the highest number of positive responses were "effectiveness" and "satisfaction", with more than 84% of responses being "totally agree" or "agree".

A large number of participants answered to the questionnaire item related to "accessibility" (a context coverage sub characteristic) as "not applicable", which may indicate that this item is not relevant for LISs from the viewpoint of the end user.

The chart in Fig. 21 shows the scoring for each characteristic using AdEQUATE score of questionnaire items (Fig. 17.b). With a median over 50, a positive perception of the quality of the LACEN system can be stated.

Figure 21 – AdEQUATE score of questionnaire items – LACEN responses



Fonte: Author's creation

5.2 DATATOX CASE STUDY

Clinical toxicology concerns the investigation, diagnosis and management of suspected poisoning cases (THOMAS, 2012), contributing to the reduction of rates related to drug abuse (NAMERA et al., 2013) and in helping victims of poisoning by several agents, such as drugs, pesticides, bites and envenomations, plants, and chemicals (MOWRY et al., 2015).

In this context, teletoxicology systems are information systems that support the referred practice over distance (CUMMINS et al., 2012; MADDRY et al., 2014; Poison Control Centers, 2015; American Association of Poison Control Centers, 2015; New Zeland National Poison Center, 2015; California Poison Center, 2015; Maryland Poison Center, 2015).

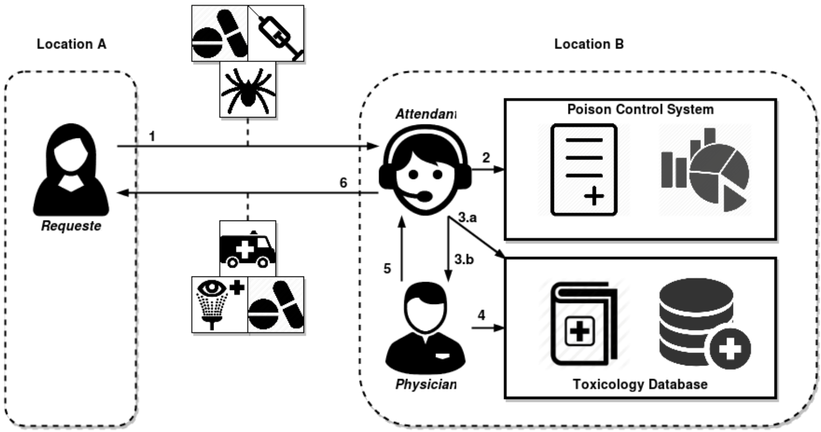
DATATOX is an example of a teletoxicology information system;

deployed in 2013, it is currently in use by 11 PCCs from 9 States in Brazil and is available 24/7. The three direct users of DATATOX include:

- patients, patient keepers, physicians as health-care requesters
- toxicologists, and other physicians as health-care providers
- telephone attendants as technical collaborators

Since its adoption, DATATOX has managed over 81,000 occurrences of poisoning caused by drugs, pesticides (both domestic and industrial), rat poison, household cleaning products, cosmetics and toiletries, industrial products, metals, abuse drugs, plants and fungi, food, venomous, and poisonous animals. Figure 22 depicts the common workflow of a poisoning occurrence.

Figure 22 – DATATOX workflow



Fonte: Author's creation

As can be seen in Fig. 22, the requester contacts the PCC providing all information available for the case (1) (e.g. patient age, patient weight, intoxication pathway). In cases involving drugs, it is important to report what kind of drug was used and/or the quantity consumed. The attendant interacts with the poison control information system (2), assessing the severity of the case. If necessary, the attendant can search for complementary information in toxicology databases (either in hard copies or in electronic form) (3.a); alternatively, the attendant may ask for the opinion of a physician (3.b). The physician, in turn, may also consult toxicology databases giving feedback (5) to the attendant. Finally, the attendant gives feedback to the requester (6) with the appropriate recommendations (e.g. first aid to be applied, medication to be prescribed).

DATATOX offers 3 main functionalities to its users: *case registry*, *case review*, and *case follow-up*. The case registry is used by attendants to register a poisoning or informative case, that may be corrected or improved by a physician using the case review functionality. The case follow-up allows the attendants to further improve the information about the case, getting in touch with requesters if needed. The DATATOX information system does not have a toxicology database module, so on-duty attendants or physicians rely on external databases such as TOXBASE©(Toxbase, 2015) and Micromedex Solutions©(Micromedex, 2015).

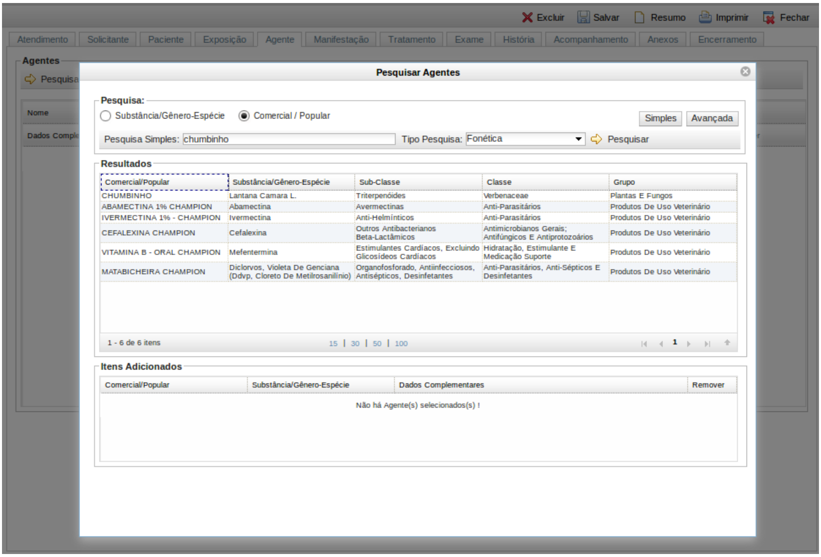
A common scenario involving DATATOX is as follows: the PCC receives a telephone call made by a requester. The call is answered by an attendant, who registers all information regarding the poisoning case within the information system and proceeds to assess the current situation of the patient. In order to support its assessment, the attendant might search external toxicology databases or ask directly to a supervising physician for information. After its assessment, the attendant gives the proper feedback to the requester, and the call is ended. If necessary, the attendant schedules a follow-up in the DATATOX information system with the requester, that might occur when a laboratory analysis arrives, or just to check the patient's evolution. The case is closed only after a careful analysis performed by a physician.

Figure 23 shows a screenshot of the system. In this screenshot, we can the system interface that enables the attendant to search for a poisoning agent.

5.2.1 Study definition

The objective of this case study is to evaluate the quality of the DATATOX system in int context of use – the public Unified Health System in Brazil. Adopting the AdEQUATE model, questionnaire items were revised and customized with counterexamples specific to toxicology systems and to the PCCs routine.

Figure 23 – Screenshot of the DATATOX System



Fonte: Author's creation

5.2.2 Study execution

In 2015 we collected data using online questionnaires available via LimeSurvey (SCHMITZ, 2017) answered by DATATOX users from multiple work shifts in all available facilities. In the main facility users were personally invited to answer the questionnaire; in other facilities, users were invited by telephone.

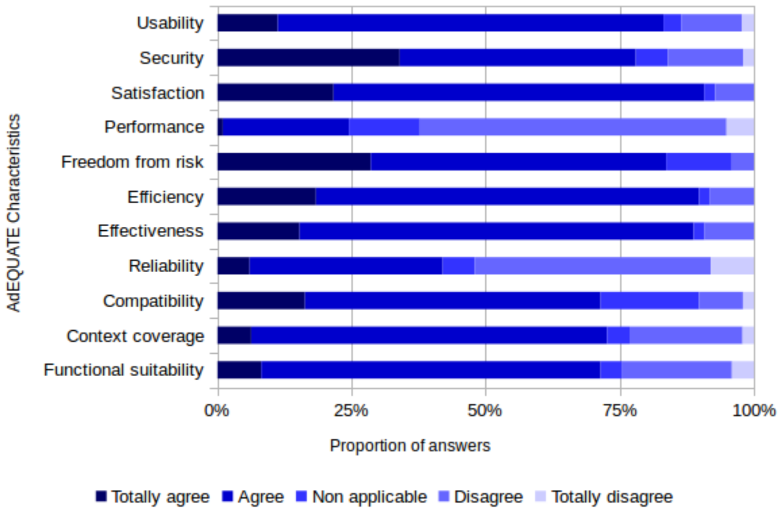
A total of 49 users were invited and answered the questionnaire. From respondents, 65% were attendants and 35% were physicians. Also, 55% of them used the system for more than a year, 31% used the system for more than six months but less than a year and only 14% of respondents used the system for less than six months.

5.2.3 Data analysis

Questionnaire items were grouped by its characteristics and using AdEQUATE metrics we have analyzed the data using descriptive statistical techniques.

The chart in Fig. 24 shows the distribution of each response option grouped by ISO/IEC 25010 quality characteristics within AdEQUATE.

Figure 24 – AdEQUATE proportion of opinion response options – DATATOX responses



Fonte: Author's creation

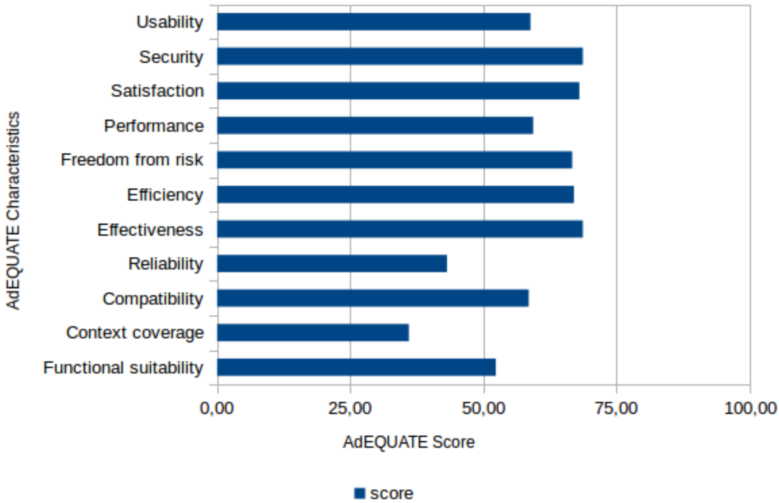
It is observed that for the majority of characteristics, the answers "totally agree" or "agree" correspond to more than 50% of the responses; exceptions are "performance", "reliability".

We can observe that "performance" have more than 50% its answers were "disagree". A strong opinion that indicates that the end-users do not perceive DATATOX as a system that is fast to interact with.

The chart depicted in Fig. 25 shows the analysis using AdEQUATE's score.

Differences between Fig. 24 and Fig. 25 show the influence of weighting a response option in the final result. While Fig. 24 shows "efficiency" and "effectiveness" roughly as having the same quality degree, Fig. 25 shows that "efficiency" has been better evaluated because it had more "totally agreed" responses. This difference is more evident in the case of "reliability" and "compatibility". "Compatibility" had more "totally agree" responses, but also had more "totally disagree" responses;

Figure 25 – AdEQUATE score of questionnaire items – DATATOX responses



Fonte: Author's creation

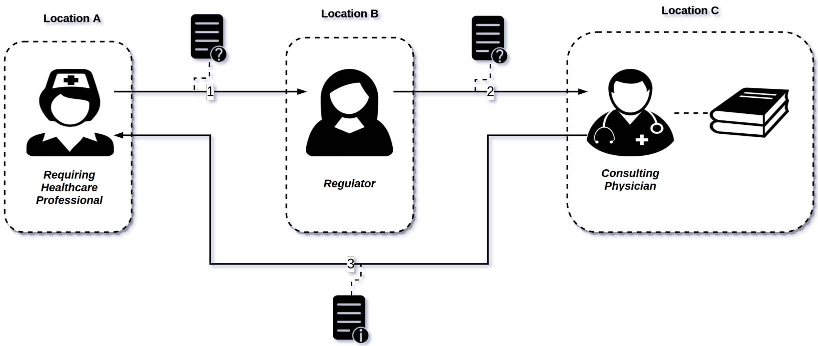
the result is that "reliability" could be perceived with a better quality degree than "compatibility". On the other hand, other characteristics were not affected by the change of metrics: "Satisfaction" and "context coverage" are still perceived as having, respectively, the highest and the lowest quality degree. Finally, since every characteristic had a score that was over two IQR away from the median, we did not identify outliers for AdEQUATE's second metric.

5.3 TELEHEALTH CASE STUDY

This case study analyses the telehealth module of the STT. Telehealth discipline, as discussed in Chap. 2, comprises a variety of nonphysician services, including, e.g., telenursing (distance monitoring of patients), telepharmacy (remote drug prescription) (WEINSTEIN et al., 2014), diabetes treatment (CHUMBLER et al., 2005), and teleconsulting (DAMASCENO et al., 2014).

The STT telehealth module has 2 main features regarding health care tele-education for health care providers that – due to geographic barriers – could not have access to a specific knowledge (CUBO et al., 2015): teleconference and teleconsulting to help health care professionals around Brazil, due to the centralization of professionals in urban areas and poor resource allocation (HADDAD et al., 2015). This case study focuses on the asynchronous teleconsulting feature whose workflow is shown in Fig. 26.

Figure 26 – STT Teleconsulting workflow



Fonte: Author's creation

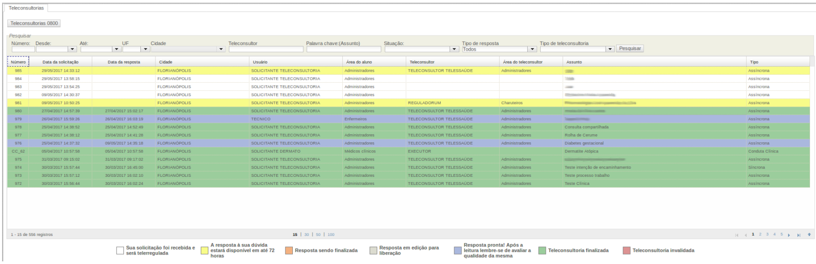
We can see that, in a location A (geographically distant from the expert physician in location C), a healthcare professional (commonly physicians and nurses) registers a healthcare related question, along with its classification (if the inquiry is either a clinical question, work process related or to forward a patient to an expert attention), its specification (the medical modality, e.g. a cardiology question, or orthopedics), and a subject (Fig. 26.1). All this information creates a teleconsulting requisition, that is stored in STT.

At another location, a regulator is responsible for analysing the teleconsulting requisition and whenever the requisition is complete and without errors, the regulator assigns the requisition to a consulting physician (Fig. 26.2).

At location C, the consulting physician may access the stored teleconsulting requisition and insert the appropriate response, comments and consulted bibliography and forwards the teleconsulting back to the requiring professional (Fig. 26.3). Optionally, the requiring professional may evaluate the answer given, providing feedback about the teleconsulting process.

Figure 27 shows a screenshot of the system. In this screenshot, we can see the system interface showing teleconsulting and their statuses.

Figure 27 – Screenshot of the Teleconsulting System



Fonte: Author's creation

5.3.1 Study definition

The objective of this case study is to evaluate the quality of the STT telehealth module, using the asynchronous teleconsulting feature in its context of use – the public Unified Health System in the State of Santa Catarina, Brazil from the perspective of the system's end-users. Through the adoption of the AdEQUATE model, questionnaire items were revised and customized with counterexamples specific to tele-education routine.

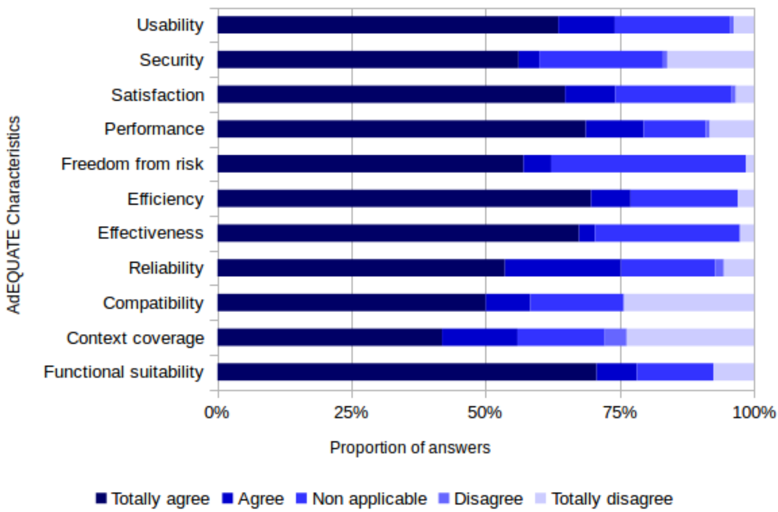
5.3.2 Study execution

During the execution phase, we applied the customized AdEQUATE questionnaire, collecting data from end users of the STT module who have been using it in their daily working routine. We had 135 responses in approximately 19% of response ratio, of which 69% respondents were requiring professionals, 22% were regulators and 9% were consulting physicians. The majority of users (43%) use the module at least once a week, followed by users that use the system in a daily basis (25%). The remainder users access the model at least twice a week (18%) and once a month (14%).

5.3.3 Data analysis

To answer the analysis question defined, questionnaire items were grouped by its characteristics and using AdEQUATE's metrics we have analyzed the data using descriptive statistical techniques, as shown in Fig. 28 and 29.

Figure 28 – AdEQUATE proportion of opinion response options – Teleconsulting responses



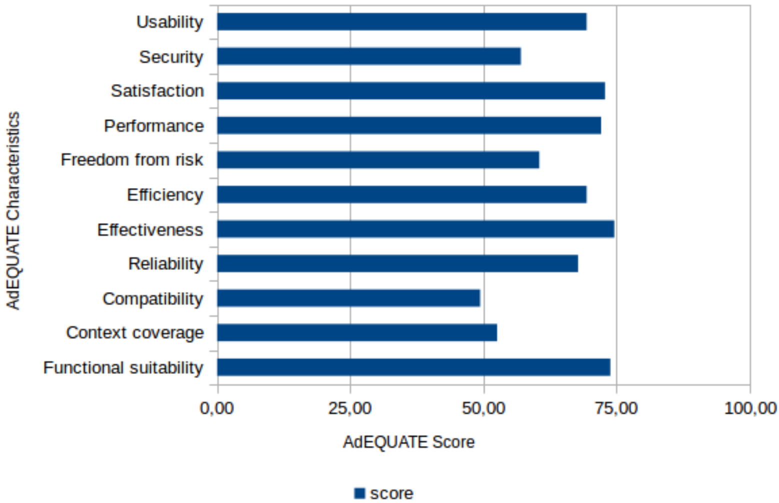
Fonte: Author's creation

The chart in Fig. 28 shows that all but "context coverage" characteristic had more than 50% of its answers as "totally agree". Regardless if we consider "totally agree" and "agree" as positive answers (because both represent degrees of high quality), all characteristics had answers greater than the 50% threshold.

The best-evaluated characteristics were "performance" and "functional suitability", in the other hand "context coverage" and "compatibility" present the best improvement opportunities.

The chart in Fig. 29 shows the scoring for each characteristic using AdEQUATE's second metric (Fig. 17.b), varying from 0 to 100. With a median over 50, a positive perception of the quality of the telehealth module can be stated.

Figure 29 – AdEQUATE score of questionnaire items – Teleconsulting responses



Fonte: Author's creation

5.4 TELEDERMATOLOGY CASE STUDY

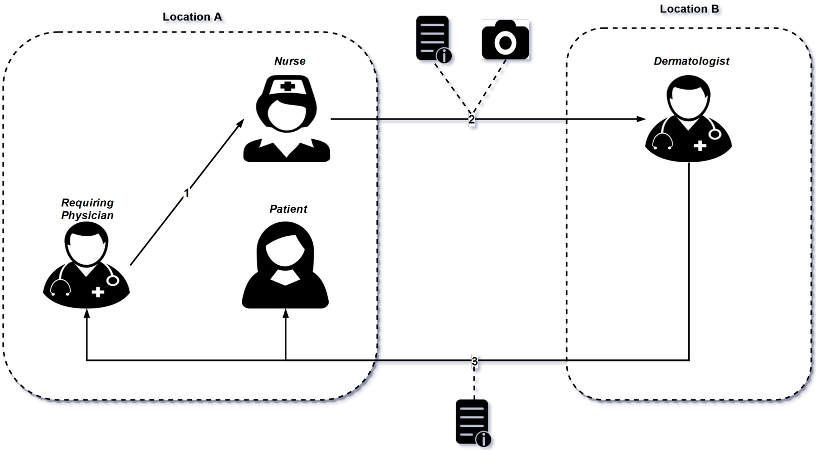
This case study analyses the teledermatology module of the STT. The teledermatology modality of telemedicine represents the exchange of dermatological information between 2 (or more) geographically separated locations, using information a communication technology, aiming for the health promotion, and education of patients, paramedics, and other healthcare professionals (MIOT; PAIXAO; WEN, 2005).

In this context, the use of teledermatology in medically underserved populations can assist in maximizing the number of patients with access to specialist care. This is largely accomplished by tele-triaging, or using telemedicine to prioritize in-person clinic visits for patients with the most urgent complaints associated with greater morbidity or mortality (LEAVITT et al., 2016).

The STT has an asynchronous teledermatology module that is used in all Santa Catarina state in public healthcare and its common workflow is depicted in Fig. 30.

In a regular consultation in the primary public health care, the physician in location A might request a teledermatology examination

Figure 30 – STT Teledermatology workflow



Fonte: Author's creation

(Fig. 30.1) – commonly because there is no dermatologist in the location. In this request, all medical data from the patient is registered in the STT, such as age, clinical condition, and the location of the lesion in the patient's anatomy.

Later, a nurse or healthcare technician proceeds to photographically register clinical relevant points of the patient's anatomy (Fig. 30.2), using a digital camera and a dermatoscope (a device for skin analysis). Once again this information is stored by the STT.

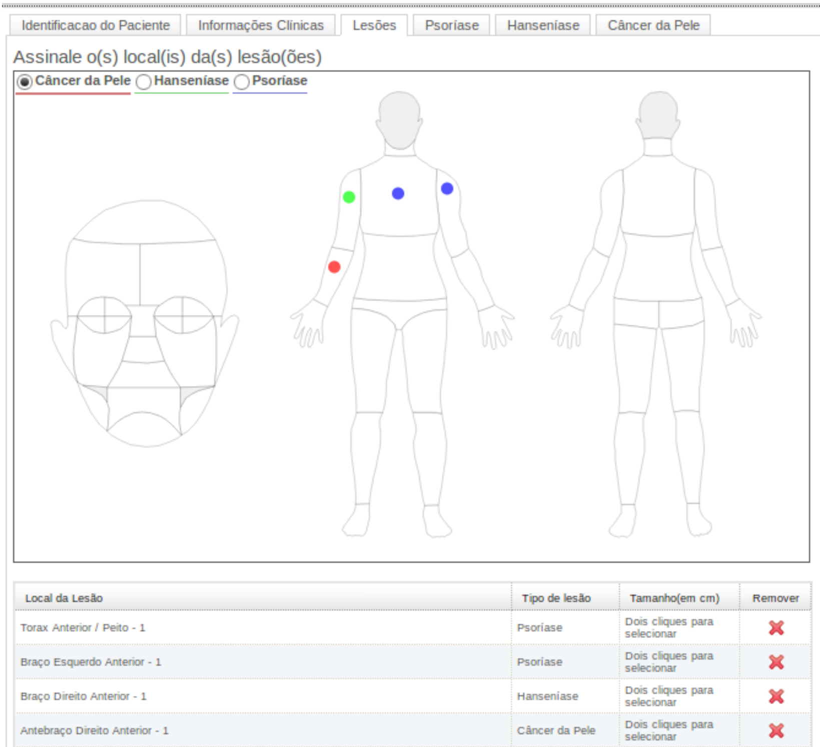
At location B, the telediagnostic physician is able to analyze the images and provide a medical report that can be accessed by both patients and requiring physician (Fig. 30.3). Furthermore, in the STT, the physician specifies the severity of the illness, indicating if the patient should be forwarded to a presidential consultation with a specialist, or milder cases the course of action to be taken by the requiring physician.

Figure 31 shows a screenshot of the system. In this screenshot, we can see the system interface showing the location tagging of skin lesions.

5.4.1 Study definition

The objective of this case study is to evaluate the quality of the STT teledermatology module, in its context of use – the public Unified

Figure 31 – Screenshot of the teledermathology module



Fonte: Author's creation

Health System in the State of Santa Catarina, Brazil from the perspective of the end-users. Through the adoption of the AdEQUATE model, questionnaire items were revised and customized with counterexamples specific to teledermatology routine.

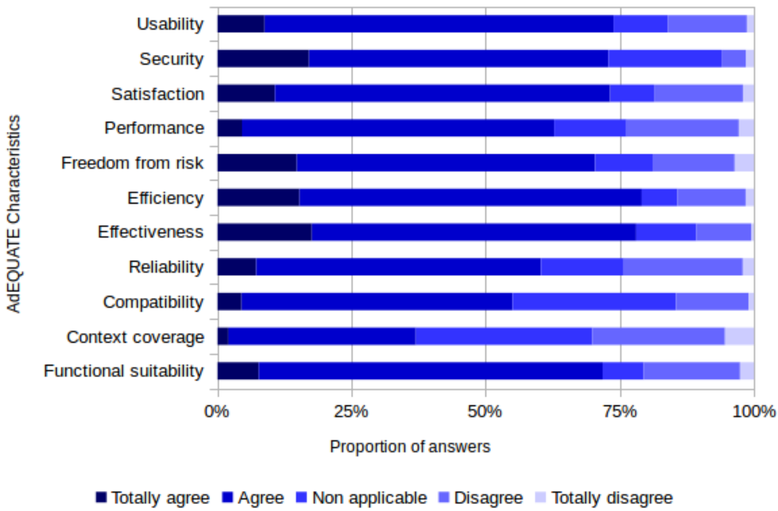
5.4.2 Study execution

During the execution phase, we applied the customized AdEQUATE questionnaire in 2016, collecting data from end users of the STT module who have been using it in their daily working routine. We had 197 responses in approximately 21% of response ratio.

5.4.3 Data analysis

Questionnaire items were grouped by its characteristics and using AdEQUATE's metrics we have analyzed the data using descriptive statistical techniques, as shown in Fig. 32 and 33.

Figure 32 – AdEQUATE proportion of opinion response options – Teledermatology responses



Fonte: Author's creation

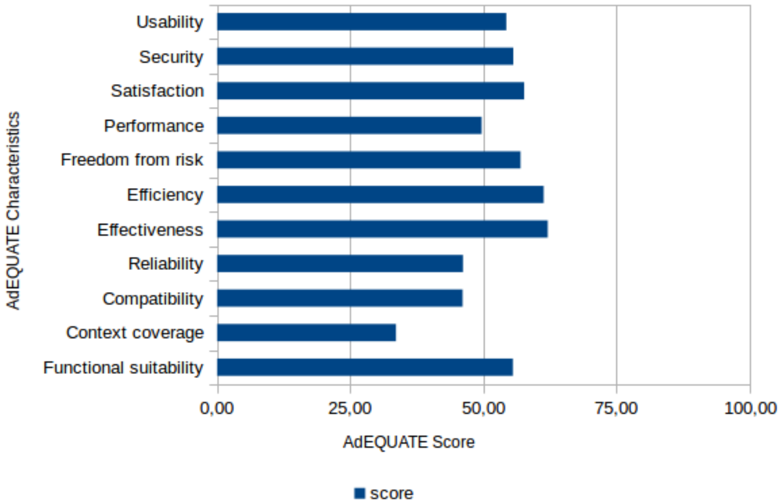
The chart in Fig. 32 shows that all characteristic present less than 25% of its answers were "totally agree". Whilst all but, once again, "context coverage" characteristic presents positive evaluations over 50%.

The "compatibility" and "context coverage" characteristics presented more than 25% of its answers being "not applicable", thus lowering the positive evaluation of those characteristics. In the other hand, efficiency and effectiveness presented the best-evaluated answers.

Using AdEQUATE's second metric, Fig. 33, we can observe that the scores fluctuate around 50, because the low amount of "totally agree" answers.

Effectiveness and efficiency present the best-evaluated characteristics, and context coverage and compatibility the poorest evaluated.

Figure 33 – AdEQUATE score of questionnaire items – Teledermatology responses



Fonte: Author's creation

5.5 TELECARDIOLOGY CASE STUDY

This case study presents the evaluation of the telecardiology module of STT. This telemedicine modality is characterized by the providing cardiovascular healthcare using information and communication technology when the distance is a critical factor (YEW et al., 2014). The common data transmitted by telecardiology is the electrocardiogram (CLARKE; JONES; LIOUPIS, 2000), that is a chart that registers electrical oscillation of cardiac tissue (KLIGFIELD et al., 2007).

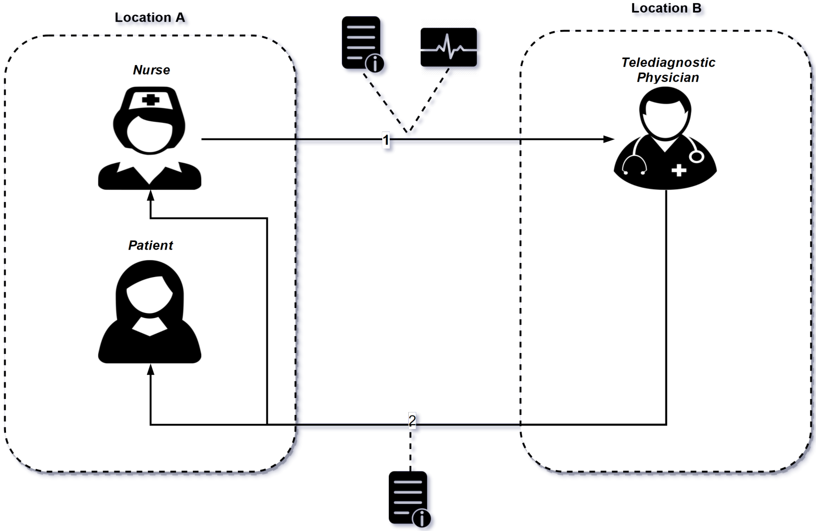
The common workflow of a telecardiology examination in the STT context is shown in Fig. 34.

In the primary health care consultation, the ECG is performed by the healthcare professional and its data along with the patient data and clinical profile is stored and send using the STT (Fig. 34.1).

At location B, the telediagnostic physician analysis the ECG and delivers a medical report using a vocabulary defined by the Brazilian Cardiology Society. This report is stored in the STT and forwarded back to the requiring professional and the patient (Fig. 34.2).

Figure 35 shows a screenshot of the system. In this screenshot,

Figure 34 – STT Telecardiology workflow



Fonte: Author's creation

we can see the system interface showing the input of healthcare risk factors that help a cardiologist interpret the ECG data.

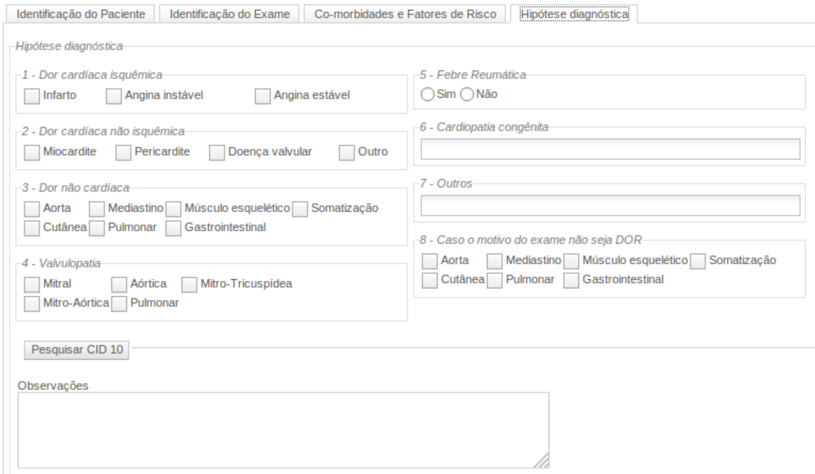
5.5.1 Study definition

The objective of this case study is to evaluate the quality of the STT telecardiology module in its context of use – the public Unified Health System in the State of Santa Catarina, Brazil from the end-users perspective. Through the adoption of the AdEQUATE model, evaluation items were revised and customized with counterexamples specific to tele-education routine.

5.5.2 Study execution

During the execution phase, we applied the customized AdEQUATE questionnaire in 2016, collecting data from end users of the STT module who have been using it in their daily working routine. We had 197 responses in approximately 26% of response ratio.

Figure 35 – Screenshot of the teledermathology module



Fonte: Author's creation

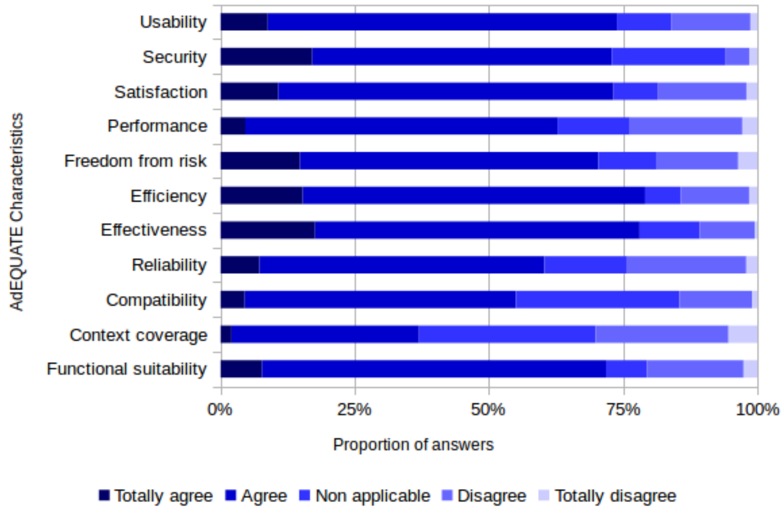
5.5.3 Data analysis

Questionnaire items were grouped by its characteristics and using AdEQUATE's metrics we have analyzed the data using descriptive statistical techniques, as shown in Fig. 36 and 37.

The chart in Fig. 36 shows that the characteristics with the highest number of "totally agree" and "agree" answers were "efficiency" and "effectiveness", with more than 75% of the answers. We can see that all other characteristics, except "context coverage", have more than 50% of responses "totally agree" and "agree". The poorest evaluated characteristic was "context coverage": less than 50% of its answers were either "totally agree" and "agree"; furthermore, this characteristic had the highest number of "disagree" answer. Characteristics "compatibility" and "context coverage" also had the highest number of "not applicable", that was over than 25%.

The chart in Fig. 37 shows the AdEQUATE score of questionnaire items, and it shows that all scores fluctuate around 50. Furthermore, as expected, "efficiency" and "effectiveness" had the highest scores, while "context coverage" and "compatibility" had the lowest.

Figure 36 – AdEQUATE proportion of opinion response options – Telecardiology responses



Fonte: Author's creation

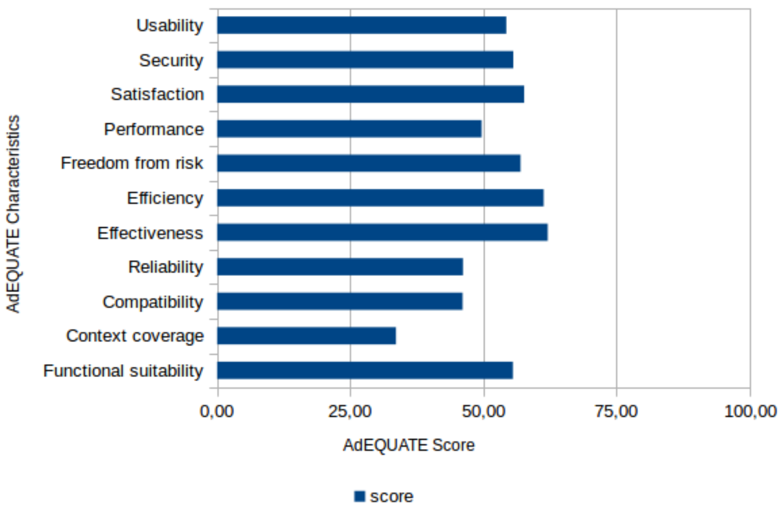
5.6 CASE STUDY DISCUSSION

A summary of the data gathered from each case study is shown in the next chart in Fig. 38. The center of the net represents the minimum score (0), while the outer rim represents the maximum score (100). This chart shows that the end-users all different case studies perceive similar levels of software quality across systems. We can also notice that, from the evaluated systems, the telehealth module of the STT is the best-evaluated systems, and the LACEN is the system that has most improvement opportunities.

Analyzing these data, we can notice a tendency towards a perception of a better quality degree of effectiveness, efficiency. This shows that those systems favor accuracy and completeness in tasks spending few resources. Likewise, this also shows a particular concern to developing systems that are precise, because telemedicine and telehealth should avoid errors, and are allocate few resources such as time.

The same improvements opportunities are noticed throughout the case studies: compatibility, and context coverage. This demonstrates a low perception of quality, as those systems are perceived as not

Figure 37 – AdEQUATE score of questionnaire items – Telecardiology responses



Fonte: Author's creation

communicating to all desired systems. Furthermore, end-users perceive a gap regarding the functionalities of those systems, i.e., to cover the complete context of the telemedicine and telehealth systems.

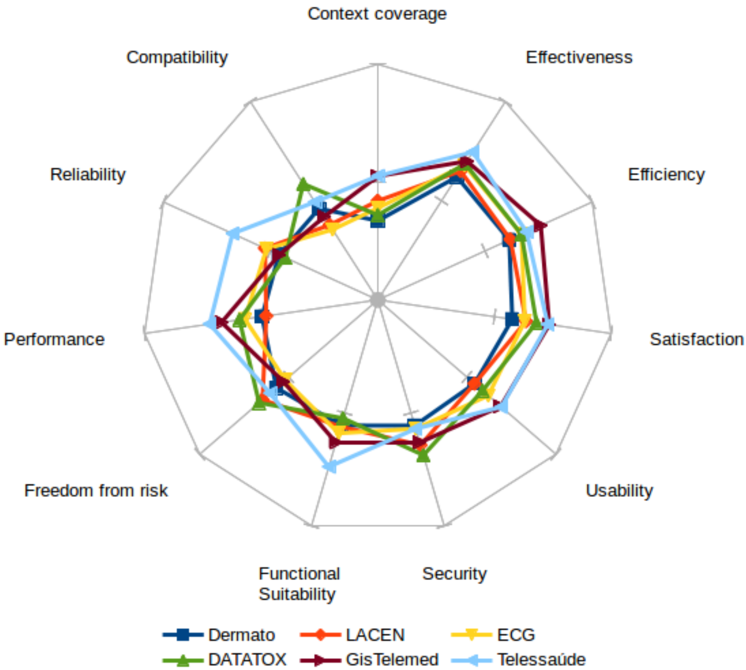
All systems analyzed in this Chapter, have been developed by the Telemedicine Laboratory[2]. Hence, the similar perception of software quality degree in most quality characteristics is comprehensible.

5.6.1 Threats to validity

As in any research, the pilot case studies are subject to threats to their validity. Since all case studies have been developed in a single stage (Chap. 1, Sec. 1.5.2), we have grouped all threats to validity in this section, that describes them and the mitigation techniques used to avoid such them.

[2] http://site.telemedicina.ufsc.br/

Figure 38 – Comparative score of all case studies



Fonte: Author's creation

5.6.1.1 Internal validity

Threats to internal validity are related to factors outside of the case study design could account for the results obtained from the evaluation (WOHLIN et al., 2012). In this context, we have to consider selection bias, mortality, and history threats.

Selection bias occurs when respondents are selected from populations that do not represent the entire group of users (WOHLIN et al., 2012). To mitigate such threat we have invited using e-mail all users that have been active (used the system or system module, in the last year). Furthermore, whenever possible we have personally invited users to respond the questionnaire, such was the LACEN, GISTelemed and DATATOX case studies. This ensures that different stakeholders have respondent the questionnaires, but also respondents with different experience levels, hence we believe that the sample of respondents is representative of users of telemedicine and telehealth systems.

The mortality effect is related to participants were excluded from the study (WOHLIN et al., 2012). In the present research have considered only complete responses, thus we are subject to mortality effect since some users have answered the questionnaire only partially. However since this dropouts respondents represent a small fraction of the answers, we do not consider this behavior as a threat to our internal validity.

5.6.1.2 Construct validity

Mono-method bias refers to measures or observations. For example, if we had defined just one measure of software quality such as, e.g., effectiveness there is a threat that the model was evaluating software quality just partially. We tried to avoid such threat by decomposing the model's goal into 11 different analysis questions, based on the ISO/IEC 25010 international standard.

5.6.1.3 External validity

With respect to construct validity, there is the bias in experimental design (WOHLIN et al., 2012): The AdEQUATE model and its measuring instrument have not been extensively validated in terms of reliability nor construct validity by a large population. We had to conduct these case studies in order to gather enough data to perform such validation, regardless we think that this threat is partially mitigated because the questionnaire items of the AdEQUATE have been defined based on established literature.

Another bias related is known as Hawthorne Effect. Participants' reactions to being studied alter their behavior and therefore the study results (WOHLIN et al., 2012).

All systems evaluated have been used in the Brazilian public healthcare. Then the external validity is limited, because it may not represent other telemedicine and telehealth context, outside public healthcare or outside Brazilian context.

6 EVALUATION OF THE ADEQUATE MODEL

It is fundamental to assess if an evaluation model and its measuring instruments truly measure what it pretends to measure (construct validity) (DEVELLIS, 2016) and if the same measuring process provides the same results (reliability) (CARMINES; ZELLER, 1979). Therefore, we present, in this chapter, an analysis of the reliability and construct validity of the AdEQUATE model. The questionnaire is analyzed based on the data collected in the case studies applying the AdEQUATE model in practice as presented in Chap. 5.

Analyzing the AdEQUATE questionnaire response options “not applicable”, “do not understand the item” and “do not know”, we have verified that these options are seldom selected by the respondents. The notable exception are two questionnaire item:

- I can use the system even having special needs (46 % of answers were “not applicable”);
- The system allows the exchange of data with other systems when necessary (18 % of answers “do not know”);

These questionnaire items have had a high percentage of answers chosen from those response options. This shows that the model must consider that users might not analyze the software quality of systems regarding accessibility, as the majority of users do not have accessibility issues. This same data shows that users have difficulty to assess if a system exchanges data with other systems.

6.1 DEFINITION OF THE EVALUATION

The goal of this evaluation is to analyze reliability and construct validity of the AdEQUATE measurement instrument from the point of view of the researchers in the context of healthcare software quality evaluations.

The reliability of a measurement instrument can be evaluated by assessing its internal consistency. Internal consistency is related to the degree in which a set of items are measuring a single quality factor,i.e., the perception of software quality. Classically, Cronbach α analysis is used to evaluate internal consistency (DEVELLIS, 2016). Cronbach α values greater than 0.7 are considered acceptable (DEVELLIS, 2016), thus, indicating internal consistency of the measurement instrument.

The construct validity of a measurement instrument is related to its ability to measure what it proposes (CARMINES; ZELLER, 1979). Convergent and discriminant validity are two subtypes of validity that composes construct validity. Convergent validity refers to the degree in which two items of a quality dimension that theoretically should be related are indeed related. On the other hand, discriminant validity verifies if concepts or measures that theoretically should not be related are indeed unrelated. To analyze convergent and discriminant validity we have calculated the item intercorrelation and item-total correlation (DEVELLIS, 2016).

Factors are unobserved variables that describe variability among observed, correlated variables. In this context, the observed variables are AdEQUATE questionnaire items, grouped in characteristics.

In this context, using the GQM methodology, this goal is decomposed into the following analysis questions:

AQ 1. Is there evidence of internal consistency in AdEQUATE's measurement instrument?

AQ 2. Is there convergent and discriminant validity in AdEQUATE's measurement instrument?

AQ 3. How many factors influence item responses in AdEQUATE's measurement instrument?

6.2 EXECUTION

In order to collect data for the evaluation, we conducted a series of 6 case studies that evaluated telemedicine and telehealth systems in different contexts using AdEQUATE model, as described in Chapter 5 and shown in Tab. 14. In order to maximize sample size, we pooled the collected data in a single sample for data analysis. The data pooling was possible due to the similarity, in terms of definitions, methods, and measurements, of the performed case studies and standardization of the collected data.

This similarity and standardization are essential for the pooling of data (KISH, 1994). In this respect, the performed studies are similar in terms of definition (the objective is to evaluate software quality), research design (case studies), and context (telemedicine or telehealth). Furthermore, all selected case studies are standardized in terms of measures (quality characteristics), data collection method (AdEQUATE measurement instrument).

Table 14 – AdEQUATE validation case studies

System (Module)	ICT healthcare domain	Sample size
LACEN System[a]	Telemedicine	69
DATATOX[b]	Telehealth	49
STT (ECG)[c]	Telemedicine	129
STT (Dermatology)	Telemedicine	197
STT (Telehealth)	Telehealth	135
GISTelemed[d]	Telehealth	35

Fonte: Author's creation

[a](ALVES et al., 2016b)
[b](ALVES et al., 2016a)
[c](OLIVEIRA, 2016)
[d](INACIO et al., 2016)

During the data analysis we have condensed the response options "Do not know", "Do not understand" and "Not applicable" into "Not applicable" as these response options meant an omission of opinion regarding the subject.

6.3 DATA ANALYSIS

In order to answer the analysis questions, we perform a statistical evaluation. The evaluation is based on the approach for the construction of measurement instruments (DEVELLIS, 2016) in alignment with procedures for the evaluation of internal consistency and construct validity of a measurement instrument (TROCHIM; DONNELLY, 2001). Data analysis was conducted using the software JASP version 0.8.0.1 (JASP Team, 2016) and data set and result are available online[1].

6.3.1 AQ 1. Is there evidence of internal consistency in AdEQUATE's measurement instrument?

Analyzing the 68 items of the AdEQUATE questionnaire (described in Section 4.1, and shown completely in Appendix A), the value of Cronbach's α is satisfactory ($\alpha = .957$).

Hence, we can perceive a strong evidence towards consistent and

[1]https://arquivos.ufsc.br/d/ce70c8e83e/

precise answers, indicating the reliability of the measurement instrument. This indicates that there exists an acceptable internal consistency, hence, we can conclude that responses between the items are consistent and precise, indicating the reliability of the questionnaire items of the AdEQUATE measurement instrument.

6.3.2 AQ 2. Is there convergent and discriminant validity in AdEQUATE's measurement instrument?

6.3.2.1 Intercorrelation of items

To analyze the intercorrelation between items, we have used a nonparametric Spearman correlation matrix. This matrix shows the Spearman's correlation coefficient, indicating the degree of correlation between a pair of questionnaire items.

The complete matrix is available online[2], and in order to illustrate the principal points with respect to this analysis we present in Tab. 15 shows a snapshot the intercorrelation between items of the AdEQUATE questionnaire because the items related to AdEQUATE characteristics Reliability, Security and Functional Suitability show interesting aspects of AdEQUATE's convergent and discriminant validity.

Among the 11 AdEQUATE quality characteristics, we have observed that most pairs of items have average or high intercorrelation (coefficients are greater than .29 (COHEN, 2013)) considering all quality characteristics. However, some item pairs have presented a low correlation (e.g. QI-5, QI-9) within its characteristics. This shows that those items seem to measure another quality characteristic than its own.

Yet, as neither negative values have been detected nor a group of items that consistently did not present an intercorrelation with respect to all evaluated models this result indicate a convergent validity.

On the other hand, some items (e.g. QI-7, QI-15) have shown average or high correlation with items from another quality characteristic. Accordingly, there is no evidence of discriminant validity. Nevertheless, the lack of discriminant validity is acceptable, as, although the model is organized in 11 quality characteristics, all characteristics are also related to a unique factor, that is the perception of software quality of telemedicine and telehealth systems.

[2] `https://arquivos.ufsc.br/d/ce70c8e83e/`

Table 15 – Spearman coefficient matrix for Reliability, Security and Functional Suitability

				Reliability		
		QI-5	*QI-6*	*QI-7*	*QI-8*	*QI-9*
Reliability	QI-5	–	**.48**	**.49**	**.30**	.22
	QI-6		–	**.53**	.27	.29
	QI-7			–	**.38**	.28
	QI-8				–	.29
	QI-9					–
				Security		
		QI-10	*QI-11*	*QI-12*	*QI-13*	*QI-14*
Reliability	QI-5	.17	.15	.23	.24	.21
	QI-6	.11	.09	.17	.20	.15
	QI-7	.19	.16	.18	.21	.27
	QI-8	.16	.12	.15	.17	.16
	QI-9	.16	.16	.14	.08	.10
Security	QI-10	–	**.62**	**.34**	**.39**	**.39**
	QI-11		–	**.36**	**.40**	**.36**
	QI-12			–	**.44**	**.36**
	QI-13				–	**.61**
	QI-14					–
				Functional Suitability		
		QI-15	*QI-16*	*QI-17*	*QI-18*	*QI-19*
Reliability	QI-5	**.45**	**.42**	**.34**	**.42**	**.43**
	QI-6	**.35**	**.37**	.26	.26	.29
	QI-7	**.39**	**.36**	**.31**	**.35**	**.33**
	QI-8	.29	**.32**	**.31**	**.37**	.29
	QI-9	.23	.20	.21	.21	.21
Security	QI-10	.25	.27	.22	.25	.22
	QI-11	.23	.28	.22	.18	.16
	QI-12	**.35**	.28	.26	.25	.28
	QI-13	**.31**	.27	.25	.26	.27
	QI-14	**.30**	.27	.28	.20	.26
Functional Suitability	QI-15	–	**.47**	**.43**	**.47**	**.47**
	QI-16		–	**.49**	**.50**	**.46**
	QI-17			–	**.50**	**.39**
	QI-18				–	**.44**
	QI-19					–

Fonte: Author's creation

6.3.2.2 Item-total correlation

Each item of the instrument should have a medium or high correlation with all the other items (DEVELLIS, 2016), as this indicates that the items present consistency in comparison to the other items. Nevertheless, a low item-total correlation of an item undermines the validity of the scale, and, therefore, should be eliminated. The complete correlation table is available online[3]. The correlation coefficients between a single item of reliability, security, and functional suitability and the other items of the measurement instrument are shown in Tab. 16.

Table 16 – Item-total correlation of Reliability, Security and Functional suitability

Quality characteristic	**No. item**	**Corrected item-total correlation**	**Cronbach's α if item was deleted**
	QI-5	.523	.957
	QI-6	.356	.957
Reliability	QI-7	.434	.957
	QI-8	.386	.957
	QI-9	.356	.957
	QI-10	**.288**	.958
	QI-11	**.276**	.958
Security	QI-12	.291	.957
	QI-13	.340	.957
	QI-14	.294	.957
	QI-15	.546	.957
	QI-16	.549	.957
Functional suitability	QI-17	.503	.957
	QI-18	.584	.956
	QI-19	.568	.956

Fonte: Author's creation

We have used the corrected item-total correlation method (DEVELLIS, 2016), which compares one item with every other item of the instrument, apart from itself. The reference value, considering a satisfactory correlation, is a coefficient greater than 0.29 (COHEN, 2013). In addition, Tab. 16 also shows the Cronbach's α, if an item was deleted expecting that no item should cause a substantial increase in the Cronbach's α (DEVELLIS, 2016).

Roughly all correlations are medium to high. Only item QI-10 and QI 11 of the quality characteristic Reliability present a low item-total correlation. These results show that these items do not have a

[3] `https://arquivos.ufsc.br/f/ce70c8e83e/`

satisfactory correlation with the others dimensions and, thus, suggests that they are not measuring Reliability. Furthermore, the value of Cronbach's α, if the items were deleted shows a small increase. Hence, these items need to be revised and possibly excluded. All other items demonstrated sufficient item-total correlation and satisfactory Cronbach's α if item was deleted, thus, indicating the validity of the measured quality factors.

The degree of correlation between the items shows whether the items measure (or not) the same quality characteristic, thus indicating evidence of convergent and discriminant validity. However, these results do not determine how many quality factors underlie the set of AdEQUATE measurement instrument items. With this objective, we performed a factor analysis, answering our analysis question AQ 3.

6.3.3 AQ 3. How many factors influence item responses in AdEQUATE's measurement instrument?

Based on the original definition of the AdEQUATE measurement instrument, we assume that it is influenced by 11 underlying factors:

1. Functional suitability
2. Reliability
3. Effectiveness
4. Satisfaction
5. Performance efficiency
6. Usability
7. Compatibility
8. Health & safety risk mitigation
9. Efficiency
10. Security
11. Context coverage

In order to identify the number of quality characteristics that represent the responses of the 68 questionnaire items set of AdEQUATE's measurement instrument, we performed a factor analysis.

In order to analyze whether the AdEQUATE measurement instrument items can be submitted to factor analysis process (BROWN, 2014), we used the KMO index and Bartlett's test of sphericity. These methods indicate the adequacy of factor analysis for a specific set of items (BROWN, 2014). The KMO index measures the sampling adequacy with values between 0 and 1. An index value near 1.0 supports a factor analysis and values less than 0.5 supports otherwise (DZIUBAN; SHIRKEY, 1974). Bartlett's sphericity test also indicates whether factor analysis is appropriate with values of a significance level less than 0.05 are considered acceptable (DZIUBAN; SHIRKEY, 1974). Analyzing the set of items of the AdEQUATE measurement instrument, we obtained a KMO index of 0.96 and a significance level of 0.001. Consequently, indicating that factor analysis is appropriate to analyze the number of factors that represent the responses of the AdEQUATE measurement instrument.

To apply factor analysis, the number of factors must be defined (BROWN, 2014). In this case study, we have used the KG criterion for this decision. This criterion defines that the number of factors is equal to the number of eigenvalues – the value of the variance of the all the items which is explained by a factor (GLORFELD, 1995) – greater than 1 (GLORFELD, 1995). Thus, following the KG criterion, our results show that 6 factor should be retained.

In the context of the AdEQUATE model, this means that the responses of the measurement instrument are representing six underlying concepts. Thus, the results of the factorial analysis indicate a different decomposition than the assumed one into 11 quality characteristics as proposed in the original model.

Having identified the number of underlying factors, another issue is to determine which items are loaded into which factor. In order to identify the factor loadings of the items, a rotation method is used (BROWN, 2014). In this study, we used the Oblimin rotation. Table 17 shows a snapshot the factor loadings of the items of Reliability, Security and Functional Suitability associated with the 6 factors. The complete table is available online[4]. Values less than 0.3 have been omitted because they represent low factor loadings for this sample size (BROWN, 2014).

Analysing Tab. 17 we can observe that all items of the Reliability characteristic have high factor loadings in Factor 4, suggesting that factor 4 is related to the Reliability characteristic. However, QI-9 also had a high factor loading in Factor 3. This indicates that this particular

[4] `https://arquivos.ufsc.br/f/ce70c8e83e/`

Table 17 – Factor analysis for Reliability, Security and Functional suitability

Quality Characteristic	Questionnaire item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Reliability	QI5	-	-	-	0.771	-	-
	QI6	-	-	-	0.812	-	-
	QI7	-	-	-	0.852	-	-
	QI8	-	-	-	0.474	-	-
	QI9	-	-	0.314	0.315	-	-
Security	QI10	-	-	-	-	-	0.727
	QI11	-	-	-	-	-	0.741
	QI12	-	-	-	-	-	0.468
	QI13	-	-	-	-	-	0.616
	QI14	-	-	-	-	-	0.632
Functional Suitability	QI15	-	-	-	0.393	-	-
	QI16	-	-	-	0.382	-	-
	QI17	0.378	0.339	-	-	-	-
	QI18	-	-	-	-	-	-
	QI19	0.549	-	-	-	-	-

Fonte: Author's creation

item measures other quality characteristics other than Reliability. This behavior is shared with other questionnaire items, expressing that some items should be revised to measure more specific quality characteristics.

Notwithstanding, the security characteristic has high factor loadings in Factor 6, and this factor has not been associated with other quality characteristics, hence, suggesting that this factor represents adequately Security. The items from the characteristic Functional Suitability (Tab. 17) had the same behavior of QI-9, and furthermore presents an item, QI-18, that has no factor loading greater than 0.3 in any factor. This behavior indicates that such items, yet measuring software quality, do not measure a particular software quality characteristic.

In summary, we can observe that the results of the factor analysis partially confirm the structure of the original AdEQUATE model clearly indicating a quality factor related to Security. The results however also clearly indicate a need for restructuring not only in terms of the number of underlying quality factors but also with respect to the grouping of the items with respect to the quality characteristic factors. Especially items related to the quality factors Satisfaction and user experience need to be revised, as the results of the factor analysis indicate that they may not be related to a single quality factor as proposed in the original model.

6.4 DISCUSSION

The obtained results show sufficient evidence to consider the reliability and construct validity of AdEQUATE acceptable as a model for the evaluation of software quality. In terms of reliability (AQ1), the results of the analysis indicate a satisfactory Cronbach's α, indicating the internal consistency of the AdEQUATE measurement instrument. Thus, evidencing that the items of AdEQUATE measurement instrument are consistent and precise with respect to the evaluation of software quality of telemedicine and telehealth systems.

Regarding construct validity, with respect to convergent validity (AQ2), we identified that the items belonging to the same characteristic, in general, presented a higher correlation. Thus, we can conclude that considering all quality characteristics, there exists evidence of convergent validity. This means, that regarding the dimensions, the items seem to be measuring what the measurement instrument proposes to measure.

Concerning discriminant validity (AQ2), in general, items belonging to different quality characteristics also present a medium or high correlation. Thus, no discriminant validity could be established. However, we consider that the lack of an indication of discriminant validity is acceptable, as all dimensions refer to a single quality factor measuring software quality.

Analyzing the item-total correlation, the results also indicate a sufficient item-total correlation of most items of the AdEQUATE measurement instrument. Only item QI-10, QI-11, and QI-59, corresponding to the quality characteristic Reliability and Usability respectively. In addition, the value of Cronbach's α if these items were deleted also indicated that they are not measuring the same quality factor. Thus, these items need be revised or excluded from the AdEQUATE measurement instrument.

Based on the factor analysis (AQ3) just one factor showed, as a result, the suggestion of a characteristic as originally proposed in the AdEQUATE model. The remainder items do not clearly suggest to be a part of any factors, thus, these items need to be either re-grouped or have their formulation reviewed. Another possible reason for the observed behavior may be the free translation of software quality terms to Brazilian Portuguese from the original sources.

6.4.1 Threats to Validity

As in any research, this one is subject to threats to its validity. Therefore, we have identified potential threats and applied mitigation strategies in order to minimize their impact on our research.

6.4.1.1 Construct validity

Construct validity concerns generalizing the result of the study to the concept behind the study (WOHLIN et al., 2012). In this context, some threats are related to the research design (case study) (WOHLIN et al., 2012) such as bias in experimental design. To mitigate this threat, we defined and documented a systematic methodology for our case study. In the definition of the case study, we used the GQM approach (BASILI; CALDIERA; ROMBACH, 1994) to systematically define the objective of this study and to decompose the objective into analysis questions.

6.4.1.2 External validity

Regarding external validity, a threat to the possibility to generalize the results is related to the sample size and diversity of the data used for the evaluation (WOHLIN et al., 2012). In respect to sample size and the user representativeness bias, we have used data collected from 6 case studies, involving a population of over 600 respondents. In terms of statistical significance, this is a satisfactory sample size allowing the generation of significant results (MCCALL; RICHARDS; WALTERS, 1977).

The data has been obtained from respondents spread in different telemedicine and telehealth contexts. However, the majority of the data came from Santa Catarina where it is used more prominently, with few responses from other states of Brazil. Thus we have the need to conduct a more extensive evaluation, considering different states and systems used outside the Brazilian public healthcare, considering also different users and different functionalities.

In its respective domains, we believe that the results can be generalized to other telemedicine and telehealth systems. However, all systems are related to public health care, therefore we conclude that the external validity of this research is limited to similar contexts.

6.4.1.3 Internal validity

Another issue refers to the correct choice of statistical tests for data analysis. To minimize this threat we performed a statistical evaluation based on the approach for the construction of measurement scales as proposed by DeVellis (2016).

We have addressed the selection bias by considering all complete answers.

7 CONCLUSION

By means of a systematic literature research, we have synthesized the state-of-the-art in software quality evaluation models in the context of telemedicine and telehealth systems as perceived by end users. The 16 relevant articles described 15 different models, a small number considering the quantity of articles analyzed. The SLR also showed that most models evaluate a single characteristic and that usability is the most evaluated software quality characteristic. Notwithstanding security, compatibility, and freedom from risk are not evaluated. Most of the models use questionnaires as measuring instrument, but these instruments are mostly developed without a systematic methodology for a specific application. Regarding the ICT-based healthcare domain, although 4 articles focused on elderly healthcare, we do not conclude that there is a tendency towards a particular domain.

In this context, we have identified that the development of software quality models in the ICT-based healthcare in the past 5 years has increased, showing a growing interest in this topic. However the developed models mostly have an ad-hoc development, and mostly focus on just one specific system and just a handful of software quality characteristics.

We have defined a software quality model for telemedicine and telehealth systems, based on the ISO 25010 standard models, with a measuring instrument designed as a questionnaire. This model with its instrument could measure the degree of perceived software quality by end users of telemedicine or telehealth systems providing data to improve the system software quality.

Also, we have provided an analysis model, and a spreadsheet to operationalize data analysis of data gathered by the AdEQUATE questionnaire.

The developed model has been applied in pilot case studies to evaluate the software quality of 6 telemedicine and telehealth systems. All systems are in the context of public health care with users distributed in Santa Catarina and other Brazilian states.

Analyzing the collected data, we have obtained results that indicate that the model's instrument presents an acceptable internal consistency and convergent validity. However, we also observed that the initial structure of the AdEQUATE model has to be revised as a factorial analysis indicated a subdivision of only 6 factors, in deference to the originally proposed structure of 11 quality characteristics.

The principal scientific contributions include the synthesis of the state-of-the-art in software quality evaluations models for telemedicine and telehealth system that focus on the end-user perception. Another contribution is the AdEQUATE model itself. Differently, from the ISO25010 standard models in which is based on, it defines quality characteristics that are relevant in the healthcare domain and are perceivable by end-users. It also differs from the standard models by having a standardized questionnaire to collect evaluation data.

In the context of this research, we have published 4 articles (ALVES et al., 2016b, 2016a; BECKHAUSER et al., 2016; INÁCIO; SAVARIS; ALVES, 2016) in the context of ICT-based healthcare conferences classified as Qualis B1.

The measuring instrument developed, its analysis model and the data set of end-user opinion regarding telemedicine and telehealth systems with more than 600 data points represent this research main technological contributions.

We expect that the availability of the AdEQUATE model will facilitate the evaluation of the software quality of telemedicine and telehealth systems leading to a better, safer and healthcare solutions.

This work opens a range of new future works. Revising the model following the data analysis performed in the model validation, re-organising questionnaire items in identified factors, also revising a number of questionnaire items to propose a second and improved version of the AdEQUATE model. A revised model might include other questionnaires for AdEQUATE model could be defined in the future with different objectives that will have different audiences such as developers or administrators.

Another opportunity is to, considering the abstract definition of software quality, build a fuzzy model to analyze the data gathered to improve the interpretation of the results. It is also possible to consider an imputation approach to missing data to enlarge the pool of usable data to analysis.

REFERENCES

ADAMS, A.; COX, A. L. Questionnaires, in-depth interviews and focus groups. In: . [S.l.]: Cambridge University Press, 2008.

AGNISARMAN, S. O. et al. Lessons learned from the usability assessment of home-based telemedicine systems. *Applied Ergonomics*, Elsevier, v. 58, p. 424–434, 2017.

ALVES, J. ao M. et al. Quality evaluation of poison control information systems: A case study of the datatox system. In: IEEE. *In 2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*. [S.l.], 2016. p. 30–35.

ALVES, J. ao M. et al. Software quality evaluation of the laboratory information system used in the santa catarina state integrated telemedicine and telehealth system. In: IEEE. *In 2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*. [S.l.], 2016. p. 76–81.

ALVES, J. M. et al. *AdEQUATE Software Quality Evaluation Model v1.0*. [S.l.], 2015.

ALVES, J. M. et al. Identifying and evaluating usability heuristics applicable to clinical laboratory systems. In: IEEE. *2014 IEEE 27th International Symposium on Computer-Based Medical Systems (CBMS)*. [S.l.], 2014. p. 529–530.

American Association of Poison Control Centers. *NPDS Elements*. 2015. Date last accessed 2015-12-14. <http://www.aapcc.org/data-system/npds-elements/>.

American Telemedicine Association. What is telemedicine. *URL: http://www. americantelemed. org/about-telemedicine/what-is-telemedicine [accessed 2014-03-31][WebCite Cache]*, 2013.

American Telemedicine Association. *Telemedicine Practice Guidelines*. 2017. <http://hub.americantelemed.org/resources/telemedicine-practice-guidelines>.

ANDARGOLI, A. E. et al. Health information systems evaluation frameworks: A systematic review. *International Journal of Medical Informatics*, Elsevier, v. 97, p. 195–209, 2017.

BANO, M.; ZOWGHI, D. A systematic review on the relationship between user involvement and system success. *Information and Software Technology*, Elsevier, v. 58, p. 148–169, 2015.

BASHSHUR, R. et al. The taxonomy of telemedicine. *Telemedicine and e-Health*, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 17, n. 6, p. 484–494, 2011.

BASHSHUR, R. L. Where we are in telemedicine/telehealth, and where we go from here. *Telemedicine Journal and e-Health*, Mary Ann Liebert, Inc., v. 7, n. 4, p. 273–277, 2001.

BASHSHUR, R. L.; REARDON, T. G.; SHANNON, G. W. Telemedicine: a new health care delivery system. *Annual Review of Public Health*, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 21, n. 1, p. 613–637, 2000.

BASILI, V. R.; CALDIERA, G.; ROMBACH, H. D. Goal question metric approach. *Encyclopedia of software engineering*, Wiley Online Library, 1994.

BECKHAUSER, E. et al. Are single-board computers an option for a low-cost multimodal telemedicine platform?: First tests in the context of santa catarina state integrated telemedicine and telehealth system. In: IEEE. *Computer-Based Medical Systems (CBMS), 2016 IEEE 29th International Symposium on.* [S.l.], 2016. p. 163–168.

BISCOGLIO, I.; MARCHETTI, E. An experiment of software quality evaluation in the audio-visual media preservation context. In: IEEE. *Quality of Information and Communications Technology (QUATIC), 2014 9th International Conference on the.* [S.l.], 2014. p. 118–123.

BOEHM, B. W.; BROWN, J. R.; LIPOW, M. Quantitative evaluation of software quality. In: IEEE COMPUTER SOCIETY PRESS. *Proceedings of the 2nd international conference on Software engineering.* [S.l.], 1976. p. 592–605.

BOLAND, M. R. et al. From expert-derived user needs to user-perceived ease of use and usefulness: A two-phase mixed-methods evaluation framework. *Journal of biomedical informatics*, Elsevier, v. 52, p. 141–150, 2014.

BOURQUE, P.; FAIRLEY, R. E. et al. *Guide to the software engineering body of knowledge (SWEBOK (R)): Version 3.0.* [S.l.]: IEEE Computer Society Press, 2014.

BREU, R.; KUNTZMANN-COMBELLES, A.; FELDERER, M. New perspectives on software quality. *IEEE Software*, IEEE, n. 1, p. 32–38, 2014. <http://dx.doi.org/10.1109/MS.2014.9>.

BRINKMANN, S. Interview. In: *Encyclopedia of Critical Psychology.* [S.l.]: Springer, 2014. p. 1008–1010.

BROOKE, J. Sus a retrospective. *Journal of Usability Studies*, v. 8, n. 2, p. 29–40, 2013.

BROOKE, J. et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, London, v. 189, n. 194, p. 4–7, 1996.

BROWN, T. A. *Confirmatory factor analysis for applied research.* [S.l.]: Guilford Publications, 2014.

BROWN, W. et al. Assessment of the health it usability evaluation model (health-ituem) for evaluating mobile health (mhealth) technology. *Journal of biomedical informatics*, Elsevier, v. 46, n. 6, p. 1080–1087, 2013.

California Poison Center. *About the California Poison Control System.* 2015. Date last accessed 2015-12-14. <http://www.calpoison.org/about/home.html>.

CARMINES, E. G.; ZELLER, R. A. *Reliability and validity assessment.* [S.l.]: Sage publications, 1979.

CERBO, A. D. et al. Narrative review of telemedicine consultation in medical practice. *Patient preference and adherence*, Dove Press, v. 9, p. 65, 2015.

CHAN, M. *World now at the start of 2009 influenza pandemic.* 2009.

CHANG, L. A psychometric evaluation of 4-point and 6-point likert-type scales in relation to reliability and validity. *Applied psychological measurement*, Sage Publications Sage CA: Thousand Oaks, CA, v. 18, n. 3, p. 205–215, 1994.

CHAU, P. Y.; HU, P. J. Examining a model of information technology acceptance by individual professionals: An exploratory study. *Journal of management information systems*, Taylor & Francis, v. 18, n. 4, p. 191–229, 2002.

CHAUDHRY, B. et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Annals of internal medicine*, Am Coll Physicians, v. 144, n. 10, p. 742–752, 2006.

CHUMBLER, N. R. et al. Evaluation of a care coordination/home-telehealth program for veterans with diabetes health services utilization and health-related quality of life. *Evaluation & the health professions*, Sage Publications, v. 28, n. 4, p. 464–478, 2005.

CIMPERMAN, M.; BRENČIČ, M. M.; TRKMAN, P. Analyzing older users? home telehealth services acceptance behavior?applying an extended utaut model. *International journal of medical informatics*, Elsevier, v. 90, p. 22–31, 2016.

CLARKE, M.; JONES, R.; LIOUPIS, D. Aidman-telecardiology over a high-speed satellite network. In: IEEE. *Computers in Cardiology 2000*. [S.l.], 2000. p. 657–660.

Clinical Laboratory Standards Institute. *CLSI AUTO3-P: Laboratory Automation: Communications with Automated Software Clinical Laboratory Systems, Instruments, Devices, and Information Systems*. 2. ed. [S.l.], 2008.

COHEN, J. Statistical power analysis for the behavioral sciences: Routledge academic. *Statistical power analysis for the behavioral sciences: Routledge Academic*, 2013.

COMPEAU, L. D.; TASHAKKORI, A.; TEDDLIE, C. *Handbook of mixed methods in social & behavioral research*. [S.l.]: JSTOR, 2003.

CÔTÉ, M.-A.; SURYN, W.; GEORGIADOU, E. In search for a widely applicable and accepted software quality model for software quality engineering. *Software Quality Journal*, Springer, v. 15, n. 4, p. 401–416, 2007.

CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *psychometrika*, Springer, v. 16, n. 3, p. 297–334, 1951. <http://dx.doi.org/10.1007/BF02310555>.

CUBO, E. et al. A parkinson's disease tele-education program for health care providers in cameroon. *Journal of the neurological sciences*, Elsevier, v. 357, n. 1, p. 285–287, 2015.

CUMMINS, M. R. et al. Building a knowledge base for health information exchange between emergency departments and poison control centers. In: IEEE. *System Science (HICSS), 2012 45th Hawaii International Conference on.* [S.l.], 2012. p. 910–914.

DAMASCENO, F. R. et al. Telehealthrs project: Supporting teleconsulting with text mining to improve continuing professional development. In: IEEE. *e-Health Networking, Applications and Services (Healthcom), 2014 IEEE 16th International Conference on.* [S.l.], 2014. p. 549–554.

DAVIS, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, p. 319–340, 1989.

DAVIS, F. D.; BAGOZZI, R. P.; WARSHAW, P. R. User acceptance of computer technology: a comparison of two theoretical models. *Management science*, INFORMS, v. 35, n. 8, p. 982–1003, 1989.

DEVELLIS, R. F. *Scale development: Theory and applications.* [S.l.]: Sage publications, 2016.

DEVINE, E. B. et al. Usability evaluation of pharmacogenomics clinical decision support aids and clinical knowledge resources in a computerized provider order entry system: a mixed methods approach. *International journal of medical informatics*, Elsevier, v. 83, n. 7, p. 473–483, 2014.

DHILLON, J. S.; WÜNSCHE, B. C.; LUTTEROTH, C. Evaluation of a social web-based telehealth system. In: ACM. *Proceedings of the 24th Australian Computer-Human Interaction Conference.* [S.l.], 2012. p. 550–553.

DHILLON, J. S.; WÜNSCHE, B. C.; LUTTEROTH, C. Evaluation of a web-based telehealth system: a preliminary investigation with seniors in new zealand. In: ACM. *Proceedings of the 13th International Conference of the NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction.* [S.l.], 2012. p. 17–24.

DÖRNYEI, Z.; TAGUCHI, T. *Questionnaires in second language research: Construction, administration, and processing.* [S.l.]: Routledge, 2009.

DROMEY, G. R. A model for software product quality. *Software Engineering, IEEE Transactions on*, IEEE, v. 21, n. 2, p. 146–162, 1995.

DUCEY, A. J. Predicting tablet computer use: An extended technology acceptance model. 2013.

DZIUBAN, C. D.; SHIRKEY, E. C. When is a correlation matrix appropriate for factor analysis? some decision rules. *Psychological bulletin*, American Psychological Association, v. 81, n. 6, p. 358, 1974.

EGEA, J. M. O.; GONZÁLEZ, M. V. R. Explaining physicians' acceptance of ehcr systems: An extension of tam with trust and risk factors. *Computers in Human Behavior*, Elsevier, v. 27, n. 1, p. 319–332, 2011.

EKELAND, A. G.; BOWES, A.; FLOTTORP, S. Effectiveness of telemedicine: a systematic review of reviews. *International journal of medical informatics*, Elsevier, v. 79, n. 11, p. 736–771, 2010.

EYSENBACH, G. What is e-health? *Journal of medical Internet research*, JMIR Publications Inc., Toronto, Canada, v. 3, n. 2, p. e20, 2001.

FOSTER, S. T.; FRANZ, C. R. On the differences between information systems' users and analysts: Managing perceptions of systems quality. *Journal of Quality Management*, Elsevier, v. 3, n. 1, p. 63–77, 1998.

FUDICKAR, S. J.; FAERBER, S.; SCHNOR, B. Kopal appointment user-interface: an evaluation with elderly. In: ACM. *Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments*. [S.l.], 2011. p. 42.

FUGGETTA, A.; NITTO, E. D. Software process. In: ACM. *Proceedings of the on Future of Software Engineering*. [S.l.], 2014. p. 1–12.

GAI, K. et al. Electronic health record error prevention approach using ontology in big data. In: IEEE. *High Performance Computing and Communications (HPCC), 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), 2015 IEEE 12th International Conferen on Embedded Software and Systems (ICESS), 2015 IEEE 17th International Conference on*. [S.l.], 2015. p. 752–757.

GARLAND, R. The mid-point on a rating scale: Is it desirable. *Marketing bulletin*, v. 2, n. 1, p. 66–70, 1991.

GLORFELD, L. W. An improvement on horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and psychological measurement*, Sage Publications Sage CA: Thousand Oaks, CA, v. 55, n. 3, p. 377–393, 1995.

GRUNDY, B. L. et al. Telemedicine in critical care: an experiment in health care delivery. *Journal of the American College of Emergency Physicians*, Elsevier, v. 6, n. 10, p. 439–444, 1977.

HADDAD, A. E. et al. Formative second opinion: Qualifying health professionals for the unified health system through the brazilian telehealth program. *Telemedicine and e-Health*, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 21, n. 2, p. 138–142, 2015.

HADJI, B.; DEGOULET, P. Information system end-user satisfaction and continuance intention: A unified modeling approach. *Journal of biomedical informatics*, Elsevier, v. 61, p. 185–193, 2016.

HAILEY, D.; ROINE, R.; OHINMAA, A. Systematic review of evidence for the benefits of telemedicine. *Journal of telemedicine and telecare*, SAGE Publications Sage UK: London, England, v. 8, n. 1_suppl, p. 1–7, 2002.

HE, J.; KING, W. R. The role of user participation in information systems development: implications from a meta-analysis. *Journal of Management Information Systems*, Taylor & Francis, v. 25, n. 1, p. 301–331, 2008.

HESSE, B. W.; SHNEIDERMAN, B. ehealth research from the user?s perspective. *American journal of preventive medicine*, Elsevier, v. 32, n. 5, p. S97–S103, 2007.

HORNER, V.; COLEMAN, A. Evaluation of a guidelines-based e-health decision support system for primary health care in south africa. In: ACM. *Proceedings of the Southern African Institute for Computer Scientist and Information Technologists Annual Conference 2014 on SAICSIT 2014 Empowered by Technology*. [S.l.], 2014. p. 142.

INACIO, A. d. S. *Estratégia de recuperação e análise de informações epidemiológicas com visualização georreferenciada sobre dados médicos heterogêneos*. Dissertação (Master thesis) — Universidade Federal de Santa Catarina, 2016.

INACIO, A. de S. et al. Designing an information retrieval system for the stt/sc. In: IEEE. *e-Health Networking, Applications and Services (Healthcom), 2014 IEEE 16th International Conference on*. [S.l.], 2014. p. 500–505.

INÁCIO, A. de S.; SAVARIS, A.; ALVES, A. v. W. João M. Sistema integrado de análise e visualização epidemiológica georreferenciada online para telemedicina. In: SBC. *XXXVI Congresso da Sociedade Brasileira de Computação, XVI WIM - Workshop de Informática Médica*. [S.l.], 2016.

INACIO, A. de S. et al. Gistelemed: An online-based gis approach to epidemiological analysis in telemedicine systems. In: IEEE. *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*. [S.l.], 2016. p. 5376–5379.

International Organization for Standardization. *ISO/IEC 25010: Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models*. [S.l.], 2014.

International Standard Organization. *ISO 9241-210:2010 - Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems*. 1. ed. [S.l.], 2010.

ISTEPANIAN, R. S.; LACAL, J. C. Emerging mobile communication technologies for health: some imperative notes on m-health. In: IEEE. *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*. [S.l.], 2003. v. 2, p. 1414–1416.

JANETSCHEK, G.; BARTSCH, G.; KAVOUSSI, L. R. Transcontinental interactive laparoscopic telesurgery between the united states and europe. *The Journal of urology*, Elsevier, v. 160, n. 4, p. 1413, 1998.

JASP Team. *JASP*. 2016. <https://jasp-stats.org/>.

JOHNSON, M. P.; ZHENG, K.; PADMAN, R. Modeling the longitudinality of user acceptance of technology with an evidence-adaptive clinical decision support system. *Decision Support Systems*, Elsevier, v. 57, p. 444–453, 2014.

KASUNIC, M. *Designing an effective survey*. [S.l.], 2005.

KEMERER, C. F.; PAULK, M. C. The impact of design and code reviews on software quality: An empirical study based on psp data. *IEEE transactions on software engineering*, IEEE, v. 35, n. 4, p. 534–550, 2009.

KHOMH, F. et al. Do faster releases improve software quality?: an empirical case study of mozilla firefox. In: IEEE PRESS. *Proceedings of the 9th IEEE Working Conference on Mining Software Repositories*. [S.l.], 2012. p. 179–188.

KISH, L. Multipopulation survey designs: five types with seven shared aspects. *International Statistical Review/Revue Internationale de Statistique*, JSTOR, p. 167–186, 1994.

KITCHENHAM, B. Procedures for performing systematic reviews. *Keele, UK, Keele University*, v. 33, n. 2004, p. 1–26, 2004.

KLIGFIELD, P. et al. Recommendations for the standardization and interpretation of the electrocardiogram. *Circulation*, Am Heart Assoc, v. 115, n. 10, p. 1306–1324, 2007.

KOCH, S. Home telehealth?current state and future trends. *International journal of medical informatics*, Elsevier, v. 75, n. 8, p. 565–576, 2006.

KOPANITSA, G.; VESELI, H.; YAMPOLSKY, V. Development, implementation and evaluation of an information model for archetype based user responsive medical data visualization. *Journal of biomedical informatics*, Elsevier, v. 55, p. 196–205, 2015.

KOTHARI, C. R. *Research methodology: Methods and techniques*. [S.l.]: New Age International, 2004.

KOZIOLEK, H. Goal, question, metric. In: *Dependability metrics*. [S.l.]: Springer, 2008. p. 39–42.

KUO, K.-M.; LIU, C.-F.; MA, C.-C. An investigation of the effect of nurses' technology readiness on the acceptance of mobile electronic medical record systems. *BMC medical informatics and decision making*, BioMed Central, v. 13, n. 1, p. 1, 2013.

LAMPASONA, C. et al. Software quality modeling experiences at an oil company. In: ACM. *Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement*. [S.l.], 2012. p. 243–246.

LEAVITT, E. R. et al. Teledermatology as a tool to improve access to care for medically underserved populations: A retrospective descriptive study. *Journal of the American Academy of Dermatology*, Elsevier, v. 75, n. 6, p. 1259–1261, 2016.

LEROUGE, C.; GARFIELD, M. J. Crossing the telemedicine chasm: have the us barriers to widespread adoption of telemedicine been significantly reduced? *International journal of environmental research and public health*, Multidisciplinary Digital Publishing Institute, v. 10, n. 12, p. 6472–6484, 2013.

LEROUGE, C.; GARFIELD, M. J.; COLLINS, R. W. Telemedicine: Technology mediated service relationship, encounter, or something else? *International journal of medical informatics*, Elsevier, v. 81, n. 9, p. 622–636, 2012.

LEROUGE, C.; GARFIELD, M. J.; HEVNER, A. R. Quality attributes in telemedicine video conferencing. In: IEEE. *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on.* [S.l.], 2002. p. 2050–2059.

LEROUGE, C.; HEVNER, A. R.; COLLINS, R. W. It's more than just use: An exploration of telemedicine use quality. *Decision Support Systems*, Elsevier, v. 43, n. 4, p. 1287–1304, 2007.

LEROUSE, C. et al. Telemedicine encounter quality: comparing patient and provider perspectives of a socio-technical system. In: IEEE. *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on.* [S.l.], 2004. p. 10–pp.

LEROY, G. et al. An end user evaluation of query formulation and results review tools in three medical meta-search engines. *International journal of medical informatics*, Elsevier, v. 76, n. 11, p. 780–789, 2007.

LEWIS, J. R. Ibm computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, Taylor & Francis, v. 7, n. 1, p. 57–78, 1995.

LIU, L.; MA, Q. Perceived system performance: a test of an extended technology acceptance model. *ACM SIGMIS Database*, ACM, v. 37, n. 2-3, p. 51–59, 2006.

LONGHURST, R. Semi-structured interviews and focus groups. *Key methods in geography*, p. 117–132, 2003.

MADDRY, J. K. et al. Wartime toxicology: evaluation of a military medical toxicology telemedicine consults service to assist physicians serving overseas and in combat (2005–2012). *Journal of medical toxicology*, Springer, v. 10, n. 3, p. 261–265, 2014.

MARCIN, J. P. et al. Changes in diagnosis, treatment, and clinical improvement among patients receiving telemedicine consultations. *Telemedicine Journal & e-health*, Mary Ann Liebert, Inc. 2 Madison Avenue Larchmont, NY 10538 USA, v. 11, n. 1, p. 36–43, 2005.

MARKONIS, D. et al. User-oriented evaluation of a medical image retrieval system for radiologists. *International journal of medical informatics*, Elsevier, v. 84, n. 10, p. 774–783, 2015.

Maryland Poison Center. 2015. Date last accessed 2015-12-14. <http://www.mdpoison.com/>.

MAVRIDIS, A.; KATRIOU, S.-A.; KOUMPIS, A. An evaluation framework for eu research and development e-health projects? systems. In: SPRINGER. *International Conference on Electronic Healthcare*. [S.l.], 2008. p. 9–16.

MCCALL, J. A.; RICHARDS, P. K.; WALTERS, G. F. *Factors in Software Quality: Final Report*. [S.l.]: Information Systems Programs, General Electric Company, 1977.

MEA, V. D. What is e-health (2): The death of telemedicine? *Journal of Medical Internet Research*, JMIR Publications Inc., Toronto, Canada, v. 3, n. 2, p. e22, 2001.

Micromedex. *Web Application Access*. 2015. Date last accessed 2015-12-14. <https://www.micromedexsolutions.com/home/dispatch/ssl/true>.

Ministerio da Saúde. *Coordenação Nacional do Programa Telessaúde Brasil Redes*. 2015. <http://189.28.128.100/dab/docs/portaldab/notas_tecnicas/Nota_Tecnica_Di

MIOT, H. A.; PAIXAO, M. P.; WEN, C. L. Teledermatologia-: Passado, presente e future. *Anais brasileiros de dermatologia*, Sociedade Brasileira de Dermatologia, v. 80, n. 5, p. 523–532, 2005.

MOODY, D. L. Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions. *Data & Knowledge Engineering*, Elsevier, v. 55, n. 3, p. 243–276, 2005.

MOWRY, J. et al. 2014 annual report of the american association of poison control centers' national poison data system (npds): 32nd annual report. *Clinical toxicology*, Taylor & Francis, v. 53, n. 10, 2015.

NAMERA, A. et al. Simple segmental hair analysis for α-pyrrolidinophenone-type designer drugs by monospin extraction for evaluation of abuse history. *Journal of Chromatography B*, Elsevier, v. 942, p. 15–20, 2013.

New Zeland National Poison Center. *Hints and Tips for Calling a Poison Centre*. 2015. Date last accessed 2015-12-14. <http://www.poisons.co.nz/calling.php>.

NOBRE, L. F.; WANGENHEIM, A. von. Development and implementation of a statewide telemedicine/telehealth system in the state of santa catarina, brazil. In: HO, K. et al. (Ed.). *Technology Enabled Knowledge Translation for eHealth*. [S.l.]: Springer New York, 2012, (Healthcare Delivery in the Information Age). p. 379–400. ISBN 978-1-4614-3494-8.

NOCUM, A. A. et al. Ergonomic evaluation and design of a mobile application for maternal and infant health for smartphone users among lower-income class filipinos. *Procedia Manufacturing*, Elsevier, v. 3, p. 5411–5418, 2015.

NORRIS, T. E. et al. Low-bandwidth, low-cost telemedicine consultations in rural family practice. *The Journal of the American Board of Family Practice*, Am Board Family Med, v. 15, n. 2, p. 123–127, 2002.

OH, H. et al. What is ehealth?: a systematic review of published definitions. *World Hosp Health Serv*, v. 41, n. 1, p. 32–40, 2005.

OLIVEIRA, C. D. C. d. *Avaliação da Qualidade de Sistema de Telecardiologia: Um Estudo de Caso do Sistema Integrado Catarinense de Telemedicina e Telessaúde*. Dissertação (Undergraduate thesis) — Universidade Federal de Santa Catarina, 2016.

ONA - Organização Nacional de Acreditação. *Avaliação de Organizações Prestadoras de Serviços de Laboratório Clínico*. [S.l.], 2006.

OSTOJIC, V. et al. Improving asthma control through telemedicine: a study of short-message service. *Telemedicine Journal & E-Health*, Mary Ann Liebert, Inc. 2 Madison Avenue Larchmont, NY 10538 USA, v. 11, n. 1, p. 28–35, 2005.

PARASURAMAN, A. Technology readiness index (tri) a multiple-item scale to measure readiness to embrace new technologies. *Journal of service research*, SAGE Publications, v. 2, n. 4, p. 307–320, 2000.

PAVLOPOULOS, S. et al. A novel emergency telemedicine system based on wireless communication technology-ambulance. *IEEE Transactions on Information Technology in Biomedicine*, IEEE, v. 2, n. 4, p. 261–267, 1998.

PILLONI, P. et al. Assessing the user experience design as a persuasive methodology in a real world sport application. In: ACM. *Proceedings of International Conference on Advances in Mobile Computing & Multimedia*. [S.l.], 2013. p. 387.

Poison Control Centers. 2015. Date last accessed 2015-12-14. <http://triage.webpoisoncontrol.org/>.

PPGCC. *Linhas de Pesquisa*. 2017. <http://ppgcc.posgrad.ufsc.br/linhas-de-pesquisa-2/>.

RADLEY, D. C. et al. Reduction in medication errors in hospitals due to adoption of computerized provider order entry systems. *Journal of the American Medical Informatics Association*, The Oxford University Press, v. 20, n. 3, p. 470–476, 2013.

RITCHIE, J. et al. *Qualitative research practice: A guide for social science students and researchers*. [S.l.]: Sage, 2013.

ROCHA, A. R. C. d.; MALDONADO, J. C.; WEBER, K. C. Qualidade de software. *Ana Regina, José Carlos Maldonado, Kival Weber-São Paulo: Pretice Hall*, 2001.

SÁNCHEZ-PI, N.; MOLINA, J. M. Adaptation of an evaluation system for e-health environments. In: SPRINGER. *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. [S.l.], 2010. p. 357–364.

SAURO, J. SUPR-q: a comprehensive measure of the quality of the website user experience. *Journal of Usability Studies*, Usability Professionals' Association, v. 10, n. 2, p. 68–86, 2015.

SAURO, J.; LEWIS, J. R. When designing usability questionnaires, does it hurt to be positive? In: ACM. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. [S.l.], 2011. p. 2215–2224.

SCANNELL, K.; PEREDNIA, D.; KISSMAN, H. *Telemedicine: Past, Present, Future. Current Bibliographies in Medicine*. [S.l.]: Maryland: National Library of Medicine, 1995.

SCHMITZ, C. *LimeSurvey: An Open Source survey tool*. [S.l.], 2017. <http://www.limesurvey.org>.

SEPULVEDA, J. L.; YOUNG, D. S. The ideal laboratory information system. *Arch Pathol Lab Med*, v. 137, n. 8, p. 1129–1140, 2013.

SHORTLIFFE, E. H.; CIMINO, J. J. *Biomedical informatics: computer applications in health care and biomedicine*. [S.l.]: Springer Science & Business Media, 2013.

SHOWELL, C.; NOHR, C. How should we define ehealth, and does the definition matter. *Stud Health Technol Inform*, v. 180, p. 881–884, 2012.

SINGH, H. et al. Types and origins of diagnostic errors in primary care settings. *JAMA internal medicine*, American Medical Association, v. 173, n. 6, p. 418–425, 2013.

SINGH, H.; GRABER, M. Reducing diagnostic error through medical home-based primary care reform. *Jama*, American Medical Association, v. 304, n. 4, p. 463–464, 2010.

SMITH, A. C.; GRAY, L. C. Telemedicine across the ages. *Med J Aust*, v. 190, n. 1, p. 15–9, 2009.

SOOD, S. et al. What is telemedicine? a collection of 104 peer-reviewed perspectives and theoretical underpinnings. *Telemedicine and e-Health*, Mary Ann Liebert, Inc. 2 Madison Avenue Larchmont, NY 10538 USA, v. 13, n. 5, p. 573–590, 2007.

STEWART, D. W.; SHAMDASANI, P. N. *Focus groups: Theory and practice*. [S.l.]: Sage publications, 2014.

SUBASI, Ö.; LEITNER, M.; TSCHELIGI, M. A usability and accessibility design and evaluation framework for ict services. In: SPRINGER. *Symposium of the Austrian HCI and Usability Engineering Group*. [S.l.], 2009. p. 102–110.

THALI, M. J. et al. Virtopsy, a new imaging horizon in forensic pathology: virtual autopsy by postmortem multislice computed tomography (msct) and magnetic resonance imaging (mri)-a feasibility study. *Journal of forensic sciences*, ASTM AMERICAN SOCIETY FOR TESTING AND, v. 48, n. 2, p. 386–403, 2003.

THOMAS, S. H. An agenda for uk clinical pharmacology: Developing and delivering clinical toxicology in the uk national health service. *British journal of clinical pharmacology*, Wiley Online Library, v. 73, n. 6, p. 878–883, 2012.

Toxbase. *The primary clinical toxicology database of the National Poisons Information Service*. 2015. Date last accessed 2015-12-14. <https://www.toxbase.org/>.

TROCHIM, W. M.; DONNELLY, J. P. Research methods knowledge base. Atomic Dog Pub., 2001.

UNTERKALMSTEINER, M. et al. Evaluation and measurement of software process improvement—a systematic literature review. *IEEE Transactions on Software Engineering*, IEEE, v. 38, n. 2, p. 398–424, 2012.

VENKATESH, V. Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information systems research*, Informs, v. 11, n. 4, p. 342–365, 2000.

VENKATESH, V. et al. User acceptance of information technology: Toward a unified view. *MIS quarterly*, JSTOR, p. 425–478, 2003.

VIITANEN, J. et al. National questionnaire study on clinical ict systems proofs: physicians suffer from poor usability. *International journal of medical informatics*, Elsevier, v. 80, n. 10, p. 708–725, 2011.

WAGNER, S. Quality models. In: *Software Product Quality Control*. [S.l.]: Springer Berlin Heidelberg, 2013. p. 29–89. ISBN 978-3-642-38570-4.

WALLAUER, J. et al. Creating a statewide public health record starting from a telemedicine network. *IT Professional*, v. 10, n. 2, p. 12–17, 2008.

WANGENHEIM, C. G. von et al. Tailoring software process capability/maturity models for the health domain. *Health and Technology*, Springer, v. 3, n. 1, p. 11–28, 2013.

WEINSTEIN, R. S. et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. *The American journal of medicine*, Elsevier, v. 127, n. 3, p. 183–187, 2014.

WOHLIN, C. et al. *Experimentation in software engineering.* [S.l.]: Springer Science & Business Media, 2012.

WOOTTON, R. *Telehealth in the developing world.* [S.l.]: IDRC, 2009.

YEW, H. et al. A telecardiology framework for rural area. *WSEAS*, p. 140–148, 2014.

YIN, R. K. *Case study research: Design and methods.* [S.l.]: Sage publications, 2013.

APPENDIX A – AdEQUATE Questionnaire

The complete AdEQUATE questionnaire (Tab. 18, 19) grouped by characteristic and subcharacteristic.

Table 18 – The AdQUATE questionnaire – part 1

Characteristic	Subcharacteristics	Questionnaire items
Effectiveness		- I can successfully complete my tasks in the system.
		- I can use the system with accuracy.
Efficiency		- I can use the system with efficiency.
Satisfaction	Usefulness	- Using the system improves the quality of my work.
		- Using the system offers a greater control over my work.
		- The system allows me to complete the tasks more quickly.
		- The system supports critical aspects of my job.
		- Using the system increases my productivity.
		- Using the system increases my job performance.
		- Using the system allows me to perform more work that would be possible without it.
		- Using the system facilitates me to perform my job.
		- In general, I think the system is useful in my job.
		- I would recommend this systems to mu colleagues.
		- I think I would use this system frequently.
	Trust	- The system allows a good service.
		- I felt very confident using the system.
		- The service that the system provides is predictable.
	Pleasure	- Interacting with the system is usually compensating.
	Comfort	- I feel comfortable using the system.
Freedom from risk	Health and safety risk mitigation	- Using the system in my job mitigates the risk in treatment or diagnosis for the patient.
Context coverage	Context completeness	- I can customize the system for my personal preferences/needs.
	Flexibility	- I think the system is flexible to interact with.
Functional suitability	Functional completeness	- I think that the system is consistent.
		- The system has functionalities that I would expect it had.
	Functional correctness	- The system provides the correct results with the necessary degree of precision.
	Functional appropriateness	- I think that the system has the adequate complexity.
		- The system functionalities facilitate the fulfillment of my tasks.
Performance efficiency	Time behavior	- The system responds quickly.
		- I con complete my tasks quickly.
	Resource utilization	- The system uses the available resources wisely.
	Capacity	- The resources of the system is adequate.
Compatibility	Co-existence	- The system executes its functionalities efficiently while sharing environment and resources with other products.
		- When the system shares an environment and resources with other products there isn't a negative impact on the other products.
	Interoperability	- The system allows exchange of information with other systems when necessary.

Fonte: Author's creation

Table 19 – The AdQUATE questionnaire – part 2

Characteristic	Subcharacteristics	Questionnaire items
Usability	Appropriateness recognizability	- I think that the various system function are well integrated.
	Learnability	- I needed to learn few things before that I could start to use the system. - I think that most people would learn to use this system quickly. - Learning to use the system is easy for me. - Would be easy for me to become skillful in the system's use.
	Operability	- I think the system is simple to use. - I think that is easy to make the system do what I want to do. - My interaction with the system is clear and comprehensive. - I thing that the system is easy to use. - I thing that I could use the system alone. - The terminology used in the system text is comprehensive. - The symbols and icons are intuitive. - The system simulates the way that I usually organize my work. - The tasks can be performed in a straightforward manner in this system.
	Protection from user's error	- The system protects me from making errors. - Errors messages clearly indicates me how to correct the problem. - Whenever I make a mistake, it is easy to recover from the error quickly.
	User interface aesthetics	- The interface design is attractive. - The system interface is pleasant. - All relevant information is presented in the screen. - All interface elements are well combined and harmonious. - The interface design is compatible with familiar conventions of similar systems.
	Accessibility	- I can use the system with my special needs.
Reliability	Maturity	- The systems always behaves as expected. - The system never stops unexpectedly.
	Availability	- The system is always operational and accessible when necessary.
	Fault tolerance	- When there is a problem in the system I still can perform my job.
	Recoverability	- In case of interruption of fault, the system recovers properly.
Security	Confidentiality	- I am certain that the system data are available only to authorized people.
	Integrity	- I am certain that the system blocks all unauthorized access to the program or its data.
	Non-repudiation	- I trust that if the system exhibits an information, a user has generated that information.
	Accountability	- I am certain that all actions that I made in the systems were assigned to my user.
	Authenticity	- I am certain that all that is associated to my user was really done by me.

Fonte: Author's creation

ANNEX A – ISO/IEC 25010 Quality Models

Tables 20, 21, and 22 show the definition of ISO/IEC 25010 quality models characteristics and subcharacteristics.

Table 20 – ISO/IEC 25010 Quality in use model

Characteristic	Subcharacteristic	Definition
Effectiveness		Accuracy and completeness with which users achieve specified goals.
Efficiency		Resources expended in relation to the accuracy and completeness with which users achieve goals.
Satisfaction		Degree to which user needs are satisfied when a product or system is used in a specified context of use.
	Usefulness	Degree to which a user is satisfied with their perceived achievement of pragmatic goals, including the results of use and the consequences of use.
	Trust	Degree to which a user or other stakeholder has confidence that a product or system will behave as intended.
	Pleasure	Degree to which a user obtains pleasure from fulfilling their personal needs.
	Comfort	Degree to which the user is satisfied with physical comfort.
Freedom from risk		Degree to which a product or system mitigates the potential risk to economic status, human life, health, or the environment.
	Economic risk mitigation	Degree to which a product or system mitigates the potential risk to financial status, efficient operation, commercial property, reputation or other resources in the intended contexts of use.
	Health and safety risk mitigation	Degree to which a product or system mitigates the potential risk to people in the intended contexts of use.
	Environmental risk mitigation	Degree to which a product or system mitigates the potential risk to property or the environment in the intended contexts of use.
Context coverage		Degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in both specified contexts of use and in contexts beyond those initially explicitly identified.
	Context completeness	Degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in all the specified contexts of use.
	Flexibility	Degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in contexts beyond those initially specified in the requirements.

Fonte: (International Organization for Standardization, 2014)

Table 21 – ISO/IEC 25010 Product quality model – Part 1

Characteristic	Subcharacteristic	Definition
Functional suitability		Degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions.
	Functional completeness	Degree to which the set of functions covers all the specified tasks and user objectives.
	Functional correctness	Degree to which a product or system provides the correct results with the needed degree of precision.
	Functional appropriateness	Degree to which the functions facilitate the accomplishment of specified tasks and objectives.
Performance efficiency		Performance relative to the amount of resources used under stated conditions.
	Time behavior	Degree to which the response and processing times and throughput rates of a product or system, when performing its functions meet requirements.
	Resource utilization	Degree to which the amounts and types of resources used by a product or system when performing its functions meet requirements.
	Capacity	Degree to which the maximum limits of a product or system parameter meet requirements.
Compatibility		Degree to which a product, system or component can exchange information with other products, systems or components, and/or perform its required functions, while sharing the same hardware or software environment."
	Co-existence	Degree to which a product can perform its required functions efficiently while sharing a common environment and resources with other products, without detrimental impact on any other product.
	Interoperability	Degree to which two or more systems, products or components can exchange information and use the information that has been exchanged.
Usability		Degree to which a product or system can be used by specified users to achieve specified goals with,effectiveness, efficiency and satisfaction in a specified context of use.
	Appropriateness recognizability	Degree to which users can recognize whether a product or system is appropriate for their needs.
	Learnability	Degree to which a product or system can be used by specified users to achieve specified goals of learning to use the product or system with effectiveness, efficiency, freedom from risk and satisfaction in a specified context of use.
	Operability	Degree to which a product or system has attributes that make it easy to operate and control.
	User error protection	Degree to which a system protects users against making errors.
	User interface aesthetics	Degree to which a user interface enables pleasing and satisfying interaction for the user.
	Accessibility	Degree to which a product or system can be used by people with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use.
Reliability		Degree to which a system, product or component performs specified functions under specified conditions for a specified period of time.
	Maturity	Degree to which a system meets needs for reliability under normal operation.
	Availability	Degree to which a system, product or component is operational and accessible when required for use.
	Fault tolerance	Degree to which a system, product or component operates as intended despite the presence of hardware or software faults.
	Recoverability	Degree to which, in the event of an interruption or a failure, a product or system can recover the data directly affected and re-establish the desired state of the system.
Security		Degree to which a product or system protects information and data so that persons or other products or systems have the degree of data access appropriate to their types and levels of authorization.
	Confidentiality	Degree to which a product or system ensures that data are accessible only to those authorized to have access.
	Integrity	Degree to which a system, product or component prevents unauthorized access to, or modification of, computer programs or data.
	Non-repudiation	Degree to which actions or events can be proven to have taken place, so that the events or actions cannot be repudiated later.
	Accountability	Degree to which the actions of an entity can be traced uniquely to the entity.
	Authenticity	Degree to which the identity of a subject or resource can be proved to be the one claimed.

Table 22 – ISO/IEC 25010 Product quality model – Part 2

Characteristic	Subcharacteristic	Definition
Maintainability		Degree of effectiveness and efficiency with which a product or system can be modified by the intended maintainers.
	Modularity	Degree to which a system or computer program is composed of discrete components such that a change to one component has minimal impact on other components.
	Reusability	Degree to which an asset can be used in more than one system, or in building other assets.
	Analysability	Degree of effectiveness and efficiency with which it is possible to assess the impact on a product or system of an intended change to one or more of its parts, or to diagnose a product for deficiencies or causes of failures, or to identify parts to be modified.
	Modifiability	Degree to which a product or system can be effectively and efficiently modified without introducing defects or degrading existing product quality.
	Testability	Degree of effectiveness and efficiency with which test criteria can be established for a system, product or component and tests can be performed to determine whether those criteria have been met.
Portability		Degree of effectiveness and efficiency with which a system, product or component can be transferred from one hardware, software or other operational or usage environment to another.
	Adaptability	Degree to which a product or system can effectively and efficiently be adapted for different or evolving hardware, software or other operational or usage environments.
	Installability	Degree of effectiveness and efficiency with which a product or system can be successfully installed and/or uninstalled in a specified environment.
	Replaceability	Degree to which a product can be replaced by another specified software product for the same purpose in the same environment.