UNIVERSITY OF TURKU

Development and Content Validation of Scoresheet and User Manual to Assess the Quality of Health Apps

Department of Nursing Science

Master's thesis

Author:

Kaile Kubota

Supervisors:

Anni Pakarinen PhD, MHSc, RN

Lotta Hamari PhD, MHSc, PT

Prof. Zhang Yuxia PhD, RN, FAAN

31.05.2021

Turku

Master's thesis

Subject: Quality assessment of non-medical device health apps
Author: Kaile Kubota
Title: Development and Content Validation of Scoresheet and User Manual to Assess the Quality of Health Apps
Supervisors: Anni Pakarinen PhD, MHSc, RN; Lotta Hamari PhD, MHSc, PT and Prof. Zhang Yuxia PhD, RN, FAAN
Number of pages: 83 pages
Date: 31.05.2021

Abstract

There has been a continuous growth of health application programs (apps) both in web-based and mobile platforms in recent years. However, there has been no instrument available to assess the degree of quality of these apps in Finland during the time this study has been initiated.

The aim of this study was to develop an easy-to-use and practical scoresheet and user manual to assess the degree of quality of health apps as well as validate its contents using an expert panel.

The design of the study adopted a two-stage process. The first stage entails designing the instrument in which the identification of conceptual framework, item generation and determining the structure of the instrument were performed. A comprehensive literature review was undertaken as well as the examination of applicable legislations, policies and guidelines pertaining to mHealth and digital health devices. The second stage entailed judgement wherein the scoresheet was tested for face validity with a small representative sample (n=6) of intended users and two rounds of content validation using an expert panel (n=19). The Content Validity Index (CVI) both in item and scale-levels were computed.

The result of the study yielded a total of 34 content validated items categorized into five distinctive domains – Basic Details, Health Content, Technical Properties, User-orientation, Privacy and Safety. The CVIs on item-level for all items reached a favorable score of > 0,78 on the assertions of relevance and clarity. Whilst it garnered > 0,90 on scale level based on universal agreement and average.

This study paved ways for the scoresheet and user manual to proceed with further psychometric measurement procedures such as reliability, feasibility and acceptability.

Keywords: health apps, assessment, rating scale development, content validity

Tiivistelmä

Terveydenhuollon ei-lääkinnällisten sovellusten (apps) määrä on kasvanut jatkuvasti viime vuosina sekä verkko- että mobiilialustoilla. Tämän tutkimuksen aloittamisen aikana ei Suomessa kuitenkaan ollut saatavilla instrumenttia näiden sovellusten laadun arvioimiseksi.

Tämän tutkimuksen tarkoituksena oli kehittää helppokäyttöinen ja käytännöllinen mittari ja käyttöopas terveyssovellusten laadun arvioimiseksi ja validoida sen sisältö asiantuntijapaneelin avulla.

Tutkimus toteutettiin kaksivaiheisena. Ensimmäisessä vaiheessa kehitettiin mittari eli määritettiin teoreettinen viitekehys, ja muodostettiin sen perusteella mittarin osiot ja rakenne. Vaiheessa yksi toteutettiin kattava kirjallisuuskatsaus sekä koottiin yhteen terveyssovelluksia koskeva lainsäädäntö ja viralliset ohjeistukset. Tutkimuksen toisessa vaiheessa arvioitiin kehitetyn mittarin ilmivaliditeettia (face validity) tavoiteltua käyttäjäjoukkoa edustavalla tarkoituksenmukaisella otoksella (n=6). Sisältövaliditeetin (content validity) testaamiseen rekrytoitiin asiantuntijapaneeli (n=19) ja he toteuttivat mittarille kaksi validointikierrosta. Vastauksista laskettiin Content Validity Index (CVI) sekä osioiden että mittarin tasolla (item and scale levels).

Tutkimuksen tuloksena syntyi 34 kohdan asiantuntijapaneelin validoima mittari terveyssovellusten laadun arvioimiseen. Mittarin kohdat on luokiteltu aihealueittain viiteen osa-alueeseen: perustiedot, terveyteen liittyvä sisältö, tekniset ominaisuudet, käyttäjälähtöisyys, sekä yksityisyys ja turvallisuus. Kaikkien osioiden CVI pisteet saavuttivat suotuisat lukemat > 0,78 relevanssiuden ja selkeyden osalta. Lisäksi mittari sai > 0,90 pisteet koko mittarin tasolla (universal agreement and average).

Tämä tutkimus tuotti uuden tavan arvioida terveyssovellusten laatua. Jatkossa mittarin ja sen käyttöoppaan psykometrisiä ominaisuuksia kuten reliabiliteettia sekä käyttökelpoisuutta ja hyväksyttävyyttä tulee edelleen testata.

Avainsanat: terveyssovellus, arviointi, mittarin kehitys, validiteetti

Table of contents

List of Figures

List of Tables

List of Abbreviations

Apps	Application Program
CE Marking	Conformité Européenne Marking
CVI	Content Validity Index
CVR	Content Validity Ratio
EU	European Union
FIMEA	Finnish Medicines Agency
GDPR	General Data Protection Regulation
I-CVI	Item-level Content Validity Index
IPR	Intellectual Property Rights
MDD	Medical Device Directive
mHealth	Mobile Health
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
POMPS	Percent of Maximum Possible Score
S-CVI	Scale-level Content Validity Index
S-CVI/UA	Scale-level Content Validity Index on Universal Agreement
S-CVI/Ave	Scale-level Content Validity Index Average
STM	Sosiaali- ja terveysministeriö
UAS	University of Applied Sciences
UX	User Experience
VALVIRA	Sosiaali- ja terveysalan lupa- ja valvontavirasto
WCAG	Web Content Accessibility Guidelines

1 Introduction

The continuous growth of health application programs (apps) both in web-based and mobile platforms has been prevalent in recent years (Linturi & Kuusi 2018, Hamari et al. 2020). However, there has been no instrument available to assess the degree of quality of these apps in Finland during the time this study has been initiated.

Health apps offer opportunities to augment care for various health-related conditions (Wisniewski et al. 2019). They can be used for data collection, care delivery, engaging patients and monitoring purposes (Tomlinson et al. 2013). According to Albrecht, von Jan & Pramann (2013), in order to better define health apps:

"... we would like to suggest using the definition provided by the World Health Organization (WHO) in 1946 that defined health as "a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity" (WHO, 1948). Therefore, applications (apps) that are in accordance with this definition of health – including apps that deal with wellness and fitness – can be summarized as health apps."

The issues of quality and safety are two important things that should be addressed with all health apps as it is set to become a significant source of health information and guidance not only for its intended users but also for professionals (Grundy, Wang & Bero 2016). Health instruments, apparatuses and software including apps that fall under the medical device category are bound by regulations and policies within the European Union (EU) such as the Regulation 2017/745 on the investigation and sale of all forms of medical devices for human use (European Commission 2017, 2021) and the Conformité Européenne or widely known as the CE Marking (Albrecht, Hillebrand and von Jan 2018; European Commission 2020). The CE marking is a symbol that a device meets all the harmonized legal requirements in the EU and conforms with other applicable legislations and directives for safety use in health care (EU Council 1993).

On the other hand, there are health apps and that fall under the non-medical device category, like apps for health promotion that are commonly used to supplement the education of children on health behaviors and risk preventions or apps that promote mindfulness, relaxations or simply offer health information. Non-medical device apps are those that are not used in any way to diagnose, prevent, monitor, treat nor alleviate a disease, injury or handicap (European Commission 2017).

There has been an abundance of mobile health apps and it is rapidly growing constantly. Each of these apps have its own health beneficial claims and information on achieving optimal health outcomes. There has been

works on developing and validating an assessment instrument for health apps outside the EU, but published studies are limited, and most are specific to a certain app or a particular health domain (Stoyanov et al. 2015). Otherwise, the focus is directed to aesthetics, engagement and user experience review rather than on the development, validity of health content, data security or if the apps follow evidence-based guidelines. Currently, there are no regulatory policies for safety use of health apps that fall under the non-medical device category and considering the rapid growth in the digitization of health and well-being practices, there is no instrument that users can utilize to ensure the validity of health information and safety the apps provide.

Health app development in itself is a rigorous process. It often involves a multi-professional team from different fields of expertise such as the health professional, app developer or designer and the programmers. In this field, user experience and level of acceptance are the common basis of evaluation of success of the app and consequently, as having good quality. Because user satisfaction can be apparently equated to how well the app has been designed and how well it functions. The basis is more on the aesthetic and user engagement aspects. In line with health care, this is not sufficient enough. There are several factors that need to be paid attention to in order to properly evaluate the degree of quality of the app.

Domains of health content and data security should be given importance in parallel to user-orientation and the app or game's technical properties. Each domain has several dimensions that also need to be examined and often, these were not given enough emphasis in order to develop a product that truly of good quality based on the aforementioned domains and its dimensions. Although non-medical device apps are not sufficiently regulated at present, there are still applicable laws, policies and guidelines that need to be considered such as the General Data Protection Regulation or widely known as GDPR (European Commission 2016, Finlex 2018), Privacy Code of Conduct on Mobile Health (European Commission 2018) and the W3 Recommendation (2011), to name a few. It is challenging to determine which policies and recommendations are applicable to a certain health app without a definite guideline to be followed or an assessment tool that can be utilized. This study has identified several gaps in ensuring the quality and validity of health apps. There is a lack of standard when it comes to this aspect of eHealth and mHealth (mobile health). Nevertheless, to recognize these shortcomings is a good start to come up with solutions to bridge the gap.

In the hopes to address these concerns, a scoresheet and user manual was developed through a comprehensive literature review, testing for face validity and two rounds of content validation using a panel of experts. The expert panel participants that shared their knowledge in this study comprised of a diverse group of specialists in relevant fields of health technology and research. The literature review provided a framework for the construction of the instrument and in generating the items that should be examined to assess the degree of quality of health apps. The contents of the scoresheet and user manual were validated on

the assertions of relevance and clarity using the Content Validity Index both in item and scale levels. On the scale level, universal agreement of the experts and average scores were both calculated.

The ultimate goal of this undertaking is to provide a feasible instrument to assess the apps which are being used in health care context and offer the opportunity to create course of actions in regulating and filtering the apps for use and recommendations. The results of the study paved ways to proceed with further psychometric measurement testing procedures to establish the reliability, feasibility and acceptability of the scoresheet and user manual.

2 Review of Literature

2.1 Assessment of health apps

The digitalization of health promotion and well-being practices has been rapidly growing constantly over the years (Linturi & Kuusi 2018, Hamari et al. 2020). Both web-based and mHealth (mobile health) application programs (apps) has been commonly used for interventions and management of health outcomes. They have shown an enormous potential to provide positive impact on our health (Dawson et al. 2020). Digital interventions can be used in a wide array of health care domains. One of them is to promote self-efficacy of children and has been one of the growing impactful uses of health apps because they provide engaging and innovative ways to promote health literacy and healthy behavior to children, as well as to further educate the parents (Pakarinen et al, 2018). Health apps are also used in mental health (Shang et al. 2019) such as suicide prevention and depression management (Martinengo et al. 2019), also in self-management of chronic illnesses such as diabetes (Petersen & Hempler 2017). These digital tools pose as a major source of health guidance and information to the general population as well as the health care professionals and providers (Grundy et al. 2016).

However, choosing for appropriate health apps are challenging (Dawson et al. 2020). User ratings cannot be considered as a reliable source of determining the quality of mHealth apps and traditional methods of assessment cannot cope to the rapid abundance of mHealth technology (Yasini, Beranger, Desmarias, Perex & Marchand 2016; Wisniewski, Liu & Henson 2019). Some of the important considerations that users should pay attention to is the transparency of the app's data privacy practices (Sunyaev, Dehling & Taylor 2015), and the lack of empirical evidence and testing with the actual involvement of the target users (Debong, Mayer & Kober 2019; Petersen & Hempler 2017). Most apps in the mobile app stores lack theoretical foundation and involvement of its intended users in its development as emphasized by Petersen & Hempler (2017) and concerns about the struggle to determine and adequately evaluate the apps' quality and safety still remains (Grundy, Wang & Bero 2016).

2.1.1 Quality of health content

The quality and validity of health information is a major concern in health apps (Dawson et al. 2020, Hwang et al. 2019). Information offered to users should be supported by evidence-based knowledge (Debong et al. 2019, Martinengo et al. 2019, Cheng et al. 2020,) and by official guidelines as well as validated recommendations (Sunyaev et al. 2015, Grundy et al. 2016, Shang et al. 2019). As any mobile or web-based apps can make any form of claims to users, there are just a lack of resources to confirm the validity of these

information. Literatures suggest that information on trials and tests that were actually conducted with the involvement of the stakeholders is important (Grundy et al. 2016, Debong et al. 2019, Hwang et al. 2019, Dawson et al. 2020, Wisniewski et al. 2019). All studies included in this undertaking has put emphasis on how imperative it is to ensure that all health-related information in health apps should adhere to valid, reliable and high-quality health content. Claims and declarations from developers and publishers are not enough to put full confidence on these apps; but rather to have reliable and trust-worthy resources to support its content as a whole device.

2.1.2 Involvement of users

Usability testing is a part of the development process of any application program (Debong et al. 2019). Health apps will need to require involving its intended users in the entire development process as well as in its testing procedures (Cheng et al. 2020, Grundy et al. 2016, Petersen & Hempler N 2017). Taking considerations on clear specifications of user appropriateness for a health app is crucial (Dawson et al. 2020, Martinengo et al. 2019, Petersen & Hempler 2017). It is important to take into account the sociocultural aspect of the target users in addition to the appropriateness of the age group the app is intended for, as this can threaten and potentially harm ethical and cultural values (Grundy et al. 2016). Suitability according to age group, cultural affiliation and values should be examined if a specific health app or game is appropriate for use by specific group of users. Furthermore, precautionary actions should be taken such as age restrictions for viewing or downloading the app (Dawson et al. 2020)

Additional aspects for consideration when it comes to user involvement that has been pointed out in previous studies is the transparency of any cost or monetary involvement upon using the health app (Sunyaev et al. 2015). Moreover, if the app development involves any affiliation to any organization or other companies (Martinengo et al. 2019). Though it is not a direct determinant of the degree of quality of an application program, it contributes to the consideration of user and their involvement with the app's usage. It has also been pointed out that there should be a two-way communication between the app developers and the users in which the users are able to provide feedback and provide their user experience directly to the app developers (Dawson et al. 2020, Hwang et al. 2019).

2.1.3 Technical properties of health apps

The technicality of a health app deals with the over-all functionality, aesthetics and efficiency of the app (Dawson et al. 2020). Health apps should not be a "one-size-fits-all" but rather a truly developed and catered tool for the purpose of providing beneficial health outcomes to its intended users (Debong et al. 2019). Language support according to the geographical location of its target users is also a determinant of the

degree of quality of an app (Grundy et al. 2016, Martinengo et al. 2019, Shang et al. 2019). In Finland, it is recommended that both Finnish and Swedish languages (Finlex 2003) be made available for any products or services for public use. With this consideration, it adds quality value to health apps as it does not discriminate against the user language (Martinengo et al. 2019, Shang et al. 2019). Moreover, the ability of the app to upgrade its performance through updates is also a feature that should be considered (Wisniewski et al. 2019) along with its extra capability to integrate with other smart devices (Petersen & Hempler 2017). According to Dawson et al. (2020), updates are important because it assures the users that the app aims to adhere to any upcoming guidelines or recommendations pertaining to its safe use, as well as future improvements to provide better user experience.

2.1.4 Privacy and security policy

All digital equipment should have a user privacy and data security policy (Grundy et al. 2016) that clearly presents rules and intentions pertaining to the processing of personal information and its protection (European Commission 2016, Finlex 2018). Health apps should be able to offer password and data protection features in order for users to have the confidence that their information is safe and cannot be accessed by other people in any way outside their consent (Sunyaev et al. 2015, Grundy et al. 2016, Debong et al. 2019, Dawson et al. 2020, Hwang et al. 2019, Martinengo 2019, Wisniewski et al. 2019).

2.1.5 Summary of assessing the quality of health apps

To date, several health-related apps have published studies and emphasized the benefits in improving health and well-being. Gamifications for promoting health literacy to children are becoming a trend in schools, health care organizations and in health projects. The potential benefits are seemingly endless, and several studies had been presenting evidence for its exploitation. However, the basis of recommendation for use of health apps and games is still lacking because there is no instrument that critically evaluates the important dimensions of these apps without bias. As Yasini et al. (2016) mentioned, user ratings in the app stores are not dependable measurements of quality nor validity because anyone can rate and give exceptional comments for the app and the benefits they offer. Not to mention, the possibility of such marketing strategies to attract users in downloading an app.

The decision whether which health app have good quality or validity is a challenging task. The transparency of data privacy policy and use of personal information should be examined (Sunyaev et al. 2015) in order to ensure that an app will not cause harm due to misuse, either for short-term or in the future. Health content, usability, literacy demand and practical aspects of app functionality are only some of the elements that should be rigorously evaluated (Dawson et al. 2020). Studies have claimed that there are apps being used at

present fall short on theoretical foundation and involvement of its target users in the development process. Some studies even advocate for the implementation of design thinking methodology (empathize-define-ideate-prototype-test) in order to ensure that the app creates value for users (Petersen & Hempler 2017). There are a few evaluation tools available for mobile apps in general. However, their suitability to address health apps lacks emphasis on evidence-based content and inadequate tool flexibility. For example, apps to promote mindfulness versus a health app that promote healthy behaviors. Therefore, there is a gap in the assessment of health apps that focuses on identifying content quality, evidence-based information and patient outcomes research.

Health care professionals and caregivers still face the struggle to adequately evaluate the contents and functions of health apps in order to guide users in effectively and safely supporting their health (Wisniewski et al. 2019). Digital application programs that fall under the medical device category are well-regulated with legislations and policies; whilst those that do not fall under this category are left to grow in abundance for everyone to use. As Zhang (2020) mentioned, this phenomenon warrants for a cautious approach not only by health care professionals and parents, but also by policy makers.

2.2 Recommendations, guidelines and policies on health care devices

2.2.1 Act on Medical Devices

The Act on Medical Devices (Finlex 2010) is an update to the MDD or the Medical Device Directive 93/42/EEC, also known as the Council Directive concerning medical devices, and the Regulation 2017/745 on medical devices (European Commission 2017). It aims to promote safety use of all forms of health care medical equipment, for its design, manufacturing as well as the use of accessories or complementary supplies for its operations. This act also applies to all forms of marketing and sales of equipment that falls under the medical device category. According to this act, the definition of a medical device is any instrument, apparatus, software or material, either used independently or with integration functions to other equipment that are mainly for use by human to diagnose, prevent, monitor, treat or alleviate any kind of diseases or injuries.

The act requires device manufacturers and developers to clearly label and provide information on the safety use of the equipment, its storage and any form of transport that may occur in the perceived future. The act emphasizes on the declarations of performed risk analyses of the manufacturers to determine the potential harm implications it may bring to users if not used properly. Labels and information documents that come with the device should indicate the safety use protocols and at the same time, the risk for harmful effects in

any case the device is not being used as intended by its purpose. Precautions and identification of perceived risks from using the device are mandatory to be included in the supplementing document. Furthermore, all instructions or information should be provided in Finnish and Swedish or English. Conformity declarations should be made available to users in both languages and this is examined by the authorities for confirmation of its
content.

Physical devices and digital software that fall under the medical device category have class categorizations which determine the level of policy compliance an equipment requires. Regardless of which class a medical device needs to conform to, they are inspected individually by the safety authorities for medicine and medical equipment development. In Finland, the National Supervisory of Health and Welfare or in Finnish, Sosiaali- ja terveysalan lupa- ja valvontavirasto (VALVIRA) and the Finnish Medicines Agency (FIMEA) mainly carry out the responsibilities of ensuring that any device that is classified under the medical device category conforms with safety procedures and informing the users of substantial information pertaining to its development, terms of use and risks that may possibly occur during the use of the device.

2.2.2 EU Regulation on Medical Devices

In 2017, the European Commission issued a regulation on amending its past directives on medical devices with Regulation 2017/745 on medical devices (European Commission 2017) that came into force in May 2021 across all EU member states. The law required that all devices used in health care that falls under the medical device category should bear the Conformité Européenne (CE) Marking in addition to the collection of adherences stated by the regulation. The mark signifies that the medical device has undergone and successfully passed quality control and testing to be safe for use by individuals in the European region. In this legislation, the device should be able to declare the rigorous evaluation and risk management conducted in the entire life cycle of the product – from development to market launching for consumer use. Additionally, it requires the manufacturers to provide reports on continuous testing to ensure the continuous safe use of the product. This act has laid the different classifications of devices for example, “Class I” which is the least classification type is for the non-invasive devices. It has the least quality requirement compared to other higher classes. This study does not elaborate further on the different classifications; but intends to present the rationale on the regulation being applied to devices that fall under the medical device category. Furthermore, the regulation imposes on strict supervision of notified bodies per European member states to scrutinize medical device manufacturers and thus, provides tighter control. Devices that are subjected to this regulation, are required to be uniquely registered such as having individual serial numbers per device in order to be traceable by authorities. Several requirements are demanded through this law, such as safety and performance requirement on labeling for both software and on other physical devices. Clinical evidence of

the devices' claim is imperative. Manufacturers should have conducted their own studies on their products and have produced their own results in order to be approved to be used by its target users or be put in public for consumer use.

2.2.3 GDPR and Privacy Code of Conduct

The General Data Protection Regulation or GDPR (European Commission 2016) and its Finnish version, Tietosuojalaki 1050/2018 (Finlex 2018) has laid out the guidelines on the protection of users pertaining to the collection and processing of their personal information across all EU member states. The regulation was made public in 2016 but was only fully enforced in 2018. It refers to data security as a fundamental right of every person and thus all devices regardless which category it belongs to, either for consumer use or any other purposes, should provide protection and seek confirmation of consent from the users in the collection and use of their information.

In the same year the GDPR has been strictly enforced, the European Commission had released the Privacy Code of Conduct on mobile health apps. It aims to foster trusts among mobile health (mHealth) users and provided the means to identify health apps on which app provides a better data privacy and security features on their products. The intention was to devise a competitive scheme amongst mHealth application programs and devices over each other for the benefit of the consumers. The Privacy Code of Conduct has not been approved after the assessment by the EU Commission Working Party. It was concluded that the GDPR criteria should be applied and to be deemed as the general guideline. The Privacy Code of Conduct on mobile health apps has now been serving as a practical guidance for mHealth developers on the principles of data privacy and security. The code encompasses the need for explicit acquisition of user consent pertaining to the collection and processing of their information, statement on purpose of use of information, privacy implications and the users' freedom of choice should they opt out of giving their consent in the collection and processing of their personal data. The code also addresses that app developers should put utmost consideration on the development and function of their product to use the least possible invasive use of users' personal information, most especially when it involves children. Depending on the defined age limit, parental consent and national laws or legislations should be acquired, consulted and adhered to.

2.2.4 Green Paper on mHealth

The Green Paper on mobile health (mHealth) by the European Commission (2014) has tackled the use and empowerment of mobile application programs in the context of health care. It presented aspects and issues on the use of mobile phones, monitoring devices or any other wireless equipment for use either in medical or public health practice. The objective of the Green Paper was to lay out plans on the deployment of mHealth

technology. It encourages the involvement of stakeholders in the development of mHealth technology and devices to ensure that developers are creating value for its users.

2.2.5 Accessibility of the Websites and Mobile Applications of Public Sector Bodies

In 2016, the EU Directive 2016/2102 - Accessibility of the Websites and Mobile Applications of Public Sector Bodies (European Commission 2016) laid out principles and strategies to be observed in developing or designing websites and all forms of mobile application programs including mobile games. It standardized the laws on accessibility on all EU member states. However, countries are not automatically covered by the directive as a law; but rather binds them for its adoption. There has been no information on rules of its conformity on when an EU member state should implement the directive to their legislation systems. The accessibility directive serves as a requirement for public sector agencies to make sure that their website or mobile applications are accessible to all persons including those with disabilities such as in eyesight, color blindness or the functionality for dictations or voice overs for the deaf.

2.2.6 Web Content Accessibility Guidelines

The W3C Recommendation or known as Web Content Accessibility Guidelines (WCAG) by the W3 Standards Organization (2011) complements the aforementioned directive. The guidelines provide shared standards for all content on websites to meet the various needs of individuals and institutions internationally. It serves the purpose of guiding web content developers, as a tool for web authoring as well as a tool to evaluate the web content.

2.2.7 STM digitization of health care services support plans

The Finnish Ministry of Social Affairs and Health has drawn out plans on key guidelines and conditions pertaining to the digitalization of health care services and functions of its administrative agencies and sectors for the year 2025 in 2016 (STM, 2016). The document contains the ministry’s perspective of key policies that should be given emphases by the administration about its transition on digitizing their services and the development possibilities of digital business growth in the public sector. Implementation plans however, are still on the works and updates on this undertaking will be issued-out in the coming years.

2.2.8 EU Report of the Working Group on mHealth Assessment Guidelines

Th EU Commission had published the Report of the Working Group om mHealth Assessment Guidelines in 2017 (European Commission 2017) that had raised the important topics in the commission's eHealth policy, particularly on the validity and reliability of information that mHealth solutions and interventions provide to its target users, as well as the public. The report has set out criteria for privacy, transparency, safety, reliability, validity, interoperability, technical stability, effectiveness, efficacy, efficiency, accessibility, usability, scalability, user experience (UX), user-centered design, security for patients, health care professionals, public authorities, payers of social health insurance, as well as the research and academia community. Several case studies have been presented between February 2016 and March 2017 from Andalusia, Catalonia and a European third-party mHealth vendor. Additionally, existing guidelines from France, Germany and the United Kingdom were examined during the work project. However, consensus has not been reached among the member states on drafting a directive pertaining to the quality of mHealth technology and the services they provide.

2.2.9 Summary on regulation and policy needs

Medical devices and technology that are used under the medical device category are regulated in Europe through a collection of policies and regulations. However, those that are categorized into the non-medical devices are subjected to adhere to the General Data Protection Regulation alone. Although the use and processing of user's data and its protection is being given attention at the present times, there are still health and technology domains that our governing bodies should also pay attention to in order to ensure that these non-medical devices do not impose any form of danger to users' health and well-being. Other guidelines serve the purpose as mere recommendations for these devices or technology to be followed voluntarily, at present. There have been no legislations pertaining to the quality assurance of non-medical devices for use by the public.

The consideration of the rapid growth of digitization of health care services and functions has been recognized throughout the EU member states, but there has been no common agreement on further regulating those device and technology that do not offer the same function as of the medical devices. Taking this into perspective, there is a clear gap when it comes to ensuring that eHealth and mHealth devices, functions and services are of high quality and safe for use.

2.3 Developing a rating scale and validating its contents

In developing assessment instruments that are intended for use in health care context, quantitative designs such as rating scales are commonly utilized to evaluate various domains of health care. There is a plethora of studies that demonstrate the development of such instruments and it requires a rigorous process (Nunnaly & Bernstein 1994). As observed from literature, developing an assessment instrument is more than just one phase. Methods used varied from one study to another. Earlier studies suggest that it has mainly two stages (Carmines & Zeller 1979; Nunnaly & Bernstein 1994; Vladin, Åslund & Nilsson 2015). The first stage is the instrument design wherein the content domain, item generation and construction of the instrument take place. The second stage entails judgement through a designated expert panel (Zamanzadeh, G 2015). Some studies include testing phase or piloting the developed instrument after validation (Shin & Kang 2019) in the same study. It is suggested that because there is no strict format on developing assessment instruments, developers should indicate the methods used and provide information on how the scale was developed (Lynn M R 1986).

According to Lynn (1986), the use of Likert scales in providing a quantitative quality assessment is a suitable psychometric scoring design to quantify subject responses in various forms of measurement instruments such as questionnaires and surveys. Literature varies on whether to use 5 or 7-point scale (Johns R 2010, Moors, K 2014). There are no standardized guidelines on which Likert scale format suits best for a certain type of study. Using lower than 5-point may significantly produce less accurate data, unless the purpose of the study requires to measure direction rather than the strength of the item's concept. While above 7-point may potentially produce ambiguous data as the difference of each point may not be as clear to all respondents (Thoyre et al. 2014). Nevertheless, it is important to pay attention to response burden and diminish it as much as possible when using Likert scales (Wynd, Schmidt & Schaefer 2003).

There are several designs for validating an assessment instrument or test. Face and content validities are among the types that should be considered early-on in developing an instrument according to Nunnaly & Bernstein (1994). She further mentioned that face validity enhances the practicality of use of any instrument as it induces cooperation of instrument respondents in terms of ease-of-use, comprehension, suitability of format and lay-out as well as preliminary relevance. Face validity implies that an instrument appears to be feasible and practical when administered to its intended users in a real scenario of its intended purpose.

On the other hand, content validity refers to the extent to which a domain item represents the theoretical construct of what it intends to measure (Bishop & Herron 2015, Murphy K 2005) and has a vital position in any type of questionnaire design (Abdollahpour et al. 2010). Content validity is a pre-requisite for instrument validation and allows opportunity to establish reliability in the further development stages (Davis, L 1992).

3 Study Aims

The aim of this study was to develop an easy-to-use and practical scoresheet and user manual to assess the degree of quality of health apps as well as validate its contents using an expert panel.

The study on Qvalidi 2019 checklist offered a starting point in aiming to design a quantitative assessment instrument for apps used in health context that do not fall under the medical device category. The study was initiated by the members of the Qvalidi Consortium and researchers at the University of Turku's Department of Nursing Science and the Department of Future Technologies. The results from the study concluded that developing a more practical, feasible and structured assessment instrument that supports different stages of the app development process is needed. Furthermore, a scoring system that outlines a definitive appraisal of the app's health promotion quality, validity and user information security.

4 Study Design and Methods

The design of this study adopts the design-judgement approach (Carmines & Zeller 1979, Lynn M R 1986, Murphy K. 2005, Halek 2017) to facilitate its intent and purpose. The development and content validation of the scoresheet and user manual were performed in two stages. The first stage is the instrument design in which a comprehensive literature review has been conducted to support the process in identifying the conceptual framework, item generation and structuring the instrument format. The second stage is the judgement. In this stage, face validity testing was also done to determine the potential improvements needed for the scoresheet and followed by the content validation phase using an expert panel. The Content Validity Index (CVI) both for item and scale levels were calculated. Figure 1: Study Design illustrates the process.

4.1 Stage 1: Instrument design

Literature review is essential in designing assessment instruments such as scales and questionnaires (Thoyre et al. 2014; Fredricksen et al. 2019; Oldland, Botti, Hutchinson & Redley 2020). It determines the conceptual framework of the aimed scoresheet and user manual as well as in generating the domains and items of what it intends to measure (Abdollabpour, Nejat, Nourozian & Majdazadeh 2010). The PRISMA flow diagram was utilized in conducting a systematic search of literature. Mendeley Desktop Version 1.19.4 © 2008-2019 Mendeley Ltd. was used for management of articles and deduplication process.

4.1.1 Identification of conceptual framework

To identify the conceptual framework of designing and validating the content of the scoresheet, the literature search was guided by the question as follows:

- What methods were used to develop a rating scale used in health care and how to validate its content?

Scientific articles with search terms (development AND scale AND "content validity") were extracted from PubMed, CINAHL and Science Direct databases. Journal articles has been collated without date limitations. English articles limited to nursing and Medline journals with health promotion as topics were considered for both search term groups as well as peer-reviewed, open access and the availability of full text. The selection of articles included in this study was performed on a Title/
Abstract level. Studies should have developed an instrument and performed a validation procedure. Please see Table 1: Inclusion-exclusion Criteria 1.

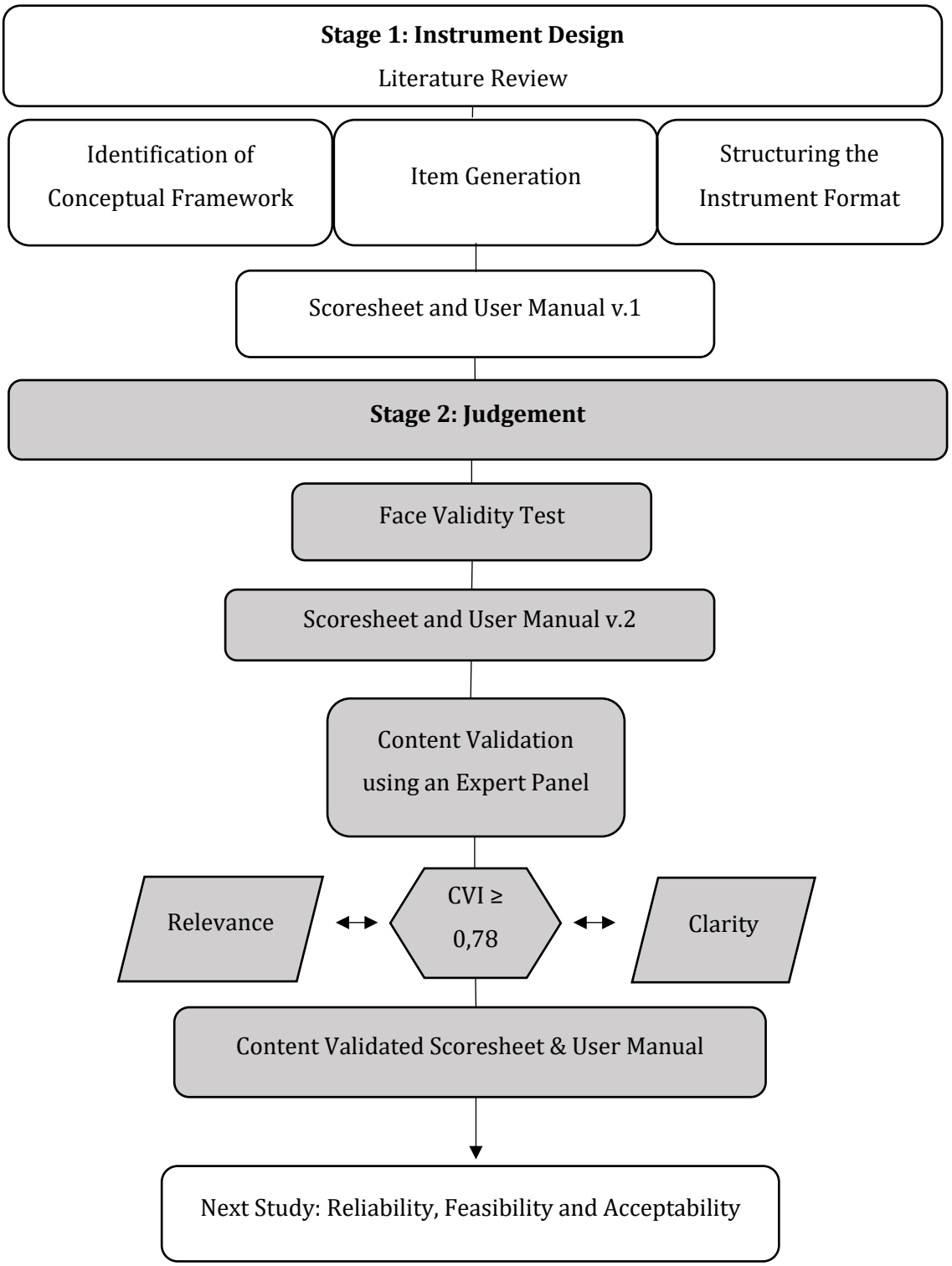

Figure 1: Study Design

Table 1: Inclusion-exclusion Criteria 1

(development AND scale AND "content validity")	Peer-reviewed Full text Open Access Nursing and Medline journals Health promotion topics Title/ Abstract English

4.1.2 Item generation

To generate the items in the scoresheet, literature has been examined by the author guided with the research question as follows:

- What factors should be considered in assessing the quality of health apps?

The articles on ("health app" OR "health game" AND assessment) were extracted from PubMed, CINAHL and Science Direct databases with time limitation of 6 years from 2014 to 2020 to focus on the most recent studies.

The consideration of applicable general guidelines and policies on the use of digital devices provided justification for the limitation. The work on mHealth Code of Conduct (European Commission 2018) initiated in April 2014 and the Commission's mHealth Green Paper in 2014 (European Commission 2014), both discussed about privacy as one of the major barriers in using apps.

Given that studies on the assessment of health apps published from this year onwards have put considerations with what has been proposed in both papers, it warrants for excluding sources from earlier than this year with regards to health app assessment. The selection of articles included in this study was also performed on a Title/ Abstract level. Please see Table 2: Inclusion-exclusion criteria.

Table 2: Inclusion-exclusion Criteria 2

(“health app” OR “health game” AND assessment)	2014 - 2020 Peer-reviewed Full text Open Access Nursing and Medline journals Health promotion topics Title/ Abstract English

The scoresheet items were generated by extracting concepts, characteristics and attributes pertaining to app quality from the articles included in the study. Basing on the research question, the words extracted were coded manually and categorized based on general concepts using Excel spreadsheet. In addition to the articles, applicable policies and guidelines referring to apps for used in health care context were examined through the Finlex Data Bank, Finnish Ministry of Social Affairs and Health (Sosiaali- ja terveysministerio/ STM), as well as the National Supervisory Authority for Welfare and Health of Finland (VALVIRA) websites.

4.1.3 Structuring the instrument format

To determine the structure of the scoresheet, the intended purpose of developing a quantifiable assessment instrument was primarily considered as well as the intended users as follows:

- Health care providers who need to assess the validity and safety of health apps intended to be used in the promotion of health and well-being prior to recommendation to users.
- Health care educators in imparting good-practice knowledge through health apps.
- Researchers who need to assess the validity and safety of different health apps for research purposes.
- App developers who need to evaluate the content and design of their prototypes.

The author recognized the potential ambiguity of technical terms and the variety of intended users. Health and technology may have terms or concepts that are not clear to the reviewer’s knowledge or it can be time consuming to check for references and clarifications. With this consideration, the user manual was developed to clarify instructions and provide further information on using the scoresheet. The definition and concepts of each item were elaborated based on literature. Additional sources such as book chapters were cited to provide further definition and examples.

4.2 Stage 2: Judgement

The judgement phase included the face validity testing of the scoresheet followed by the content validation using an expert panel. The degree of agreement across participants was calculated by percentage in the face validity phase and descriptive feedbacks based on user experience were collected; whilst the Content Validity Index in item level (I-CVI), scale level universal agreement and average (S-CVI/UA, S-CVI/Ave) were calculated in the content validation phase.

4.2.1 Testing for face validity

Measuring for face validity yields important information in developing a scale instrument (Holden 2010). It allows the developer to determine whether the items of each domain makes sense, appropriate to be included in the instrument and it also can provide the opportunity to measure the preliminary relevance of the items to the intended users (Connel et al. 2018). Boateng et. al. (2018) also suggested that it is best practice to test any instrument at the early stage as soon as possible and can help with the further development needed for the instruments. Tsang et al. (2017) also mentioned that early testing is an opportunity for the instrument developer to determine if there is any form of confusion about the items and whether respondents have suggestions for improvements. The results of the early testing can yield valuable information as potential justification in moving forward with the further development phase of the instrument.

The testing for the face validity of the initial version of the scoresheet (n=39) involved a non-medical device health app by a start-up company in Finland and a group of health technology master's students at the University of Turku (n = 6) fulfilling the criteria as shown in table 3 and are considered to have the great possibility of using the assessment tool in the future. The scoresheet was used to evaluate the app and the students were asked to fill-out an evaluation form pertaining to the use of the scoresheet. The questionnaires were distributed to the participants via Google Forms and rated the scoresheet in a 4-point Likert scale based on clarity of instruction, ease-of-use, layout and structure. In addition, participants were also asked to provide descriptive feedbacks about their experience in using the scoresheet. The test for face validity was done preserving the anonymity of the participants and consents were solicited. The purpose of the test was to determine if the scoresheet appears to be measuring what it intends to measure, as well as to identify potential improvements needed for the scoresheet based on user experience.

Table 3: Face Validity Sampling Criteria

Health app	Reviewer participants
Non-medical device app intended for use in Finland	Health and technology master's students or teaching personnel at the University of Turku who attended the presentation of the health app to be reviewed and have confirmed their participation

4.2.2 Content validation using an expert panel

The content validation of an instrument was performed based on expert judgments wherein a number of experts rated the relevance and clarity of the items on Likert scales (Davis L. 1992). Literature suggests that researchers should be critical on the criteria for the selection of expert panel participants and that it is dependent on the scope of the study (Zamanzadeh G. 2015). The selection process must be done on the basis of expert knowledge, specific training or professional experience on the subject matter (Costa A.O. 2011).

Purposive sampling was used to form the expert panel. Prospective participants (n = 19) were invited via email and asked for their confirmation to participate via webropol. The description and purpose of the study were included as an email attachment to provide information on why they received the invitation. In addition, it was assured that their participation will be maintained as anonymous and only the results of the content validation process will be presented in the study. The recruited participants should belong to any of the expertise category for at least one year in Finland. The sampling criteria are enumerated in Table 4: Expert Panel Criteria.

Table 4: Expert Panel Criteria

Expertise	Years
Games and gamification researcher, university professor	>20
Research manager, health technology innovator, university of applied science and university lecturer	>15
Health technology innovator, applied well-being technology expert, member of health care association sector	>10

University of applied science degree program director (Bachelor of Nursing, Master of Global Health Care)	>10
Public health care professional at management level / health care provider representative, digital health and well-being expert	>10
University hospital innovation agent and project specialist	>10
Principal lecturer in health care field, responsible of digital things and development of health domain	>5
Games and gamification expert	3
Registered nurse, health care and well-being researcher, focus on health technology and digital health	3

Content Validity Index (CVI) was used for the content validation of the scoresheet and user manual. It is the most widely used process in determining the validity of an instrument (Haynes et al. 1995, Kunter B. 2006, Abdollahpour et al. 2010). It provides more information on relevance as well as opportunity to improve or eliminate an item in the instrument (Polit & Beck 2006; Polit, Beck & Owens 2007). There are different forms of Content Validity Indices and in this study, item-level and both the average and universal agreement of scale-level CVIs were calculated. The indices are enumerated and defined in Table 5: Definitions of CVI Indices.

Table 5: Definitions of CVI Indices (Polit & Beck, 2006)

CVI	Degree to which an instrument has an appropriate sample of items for construct being measured.
I-CVI	Content Validity of individual items
S-CVI	Content Validity of the overall scale
S-CVI/UA	Proportion of items on a scale that were rated relevant by all experts
S-CVI/Ave	Average of the I-CVIs for all items on the scale

Each item in the scoresheet was rated on the assertions of relevance and clarity of description in the user manual in a four-point Likert-type scale (1 = Strongly disagree, 2 = Disagree, 3 = Agree and 4 = Strongly agree). The content validity index on item level (I-CVI) was calculated by adding the items with agreed ratings of three (3) or four (4) divided by the number of experts (agreed items / number of experts). Another form of content validity index is the scale level. There are two methods in calculating for the scale level content validity index – the universal agreement (S-CVI/UA) and the average of the I-CVI scores for all items on the scale (S-CVI/Ave) (Polit & Beck 2006, Bishop & Herron 2015). The CVI is expressed in terms of percentage. Please refer to table 6 for the summary of content validity index calculations.

Table 6: Content Validity Index Calculations

CVI Indices	**Calculation**
I-CVI	Sum of items rated with 3 or 4 by the experts
S-CVI/ UA	Proportion of items that achieve a relevance rating of 3 or 4 by all experts
S-SCI/ Ave	Average of I-CVI scores for all items

5 Results

The literature review conducted in this study included a total of 23 scientific articles, 11 legal statements and 11 references from book chapters to support the contents of the user manual in providing clear descriptions of items in the scoresheet. The PRISMA flow diagram was utilized in the search process and shown in Figure 2: PRISMA Flow Diagram.

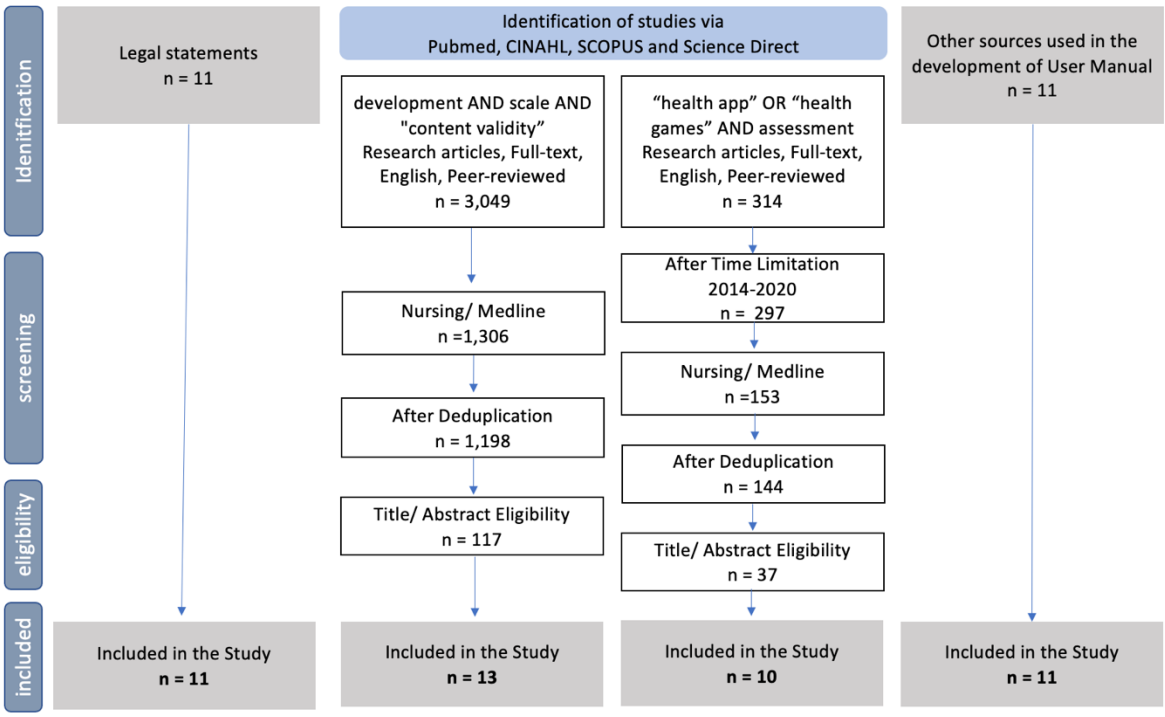

Figure 2: PRISMA Flow Diagram

The list of articles yielded from the conducted literature search in this study are listed in Table 7: Articles on development AND scale AND "content validity" and Table 8: Articles on “health app” OR “health games” AND assessment. Sources from previous study and those which were used in the user manual are listed in the references section.

Table 7: Articles on development AND scale AND "content validity

Title	Author	Year
Development and content validation of the Multifactoral assessment of perceived social support (MAPSS)	Fredericksen et al.	2019

Development and evaluation of the content validity, practicability and feasibility of the Innovative dementia-oriented Assessment system for challenging behaviour in residents with dementia	Halek, M., Holle, D., & Bartholomeyczik, S.	2017
Development and validation of Nurses' Moral Courage Scale	Numminen, O., Katajisto, J., & Leino-Kilpi, H.	2019
A Framework of Nurses' Responsibilities for Quality Health care — Exploration of Content Validity	Oldland, Botti	2020
The content validity index: are you sure you know what's being reported? Critique and recommendations	Polit, D. F., & Beck, C. T.	2006
Development and preliminary validation of the Neonatal Infant Acute Pain Assessment Scale (NIAPAS)	Pölkki, T., Korhonen, A., Axelin, A., Saarela, T., & Laukkala, H.	2014
Development and Validation of a Person-Centered Perioperative Nursing Scale	Shin, S., & Kang, J.	2019
Development, piloting and validation of the Recommending Cardiac Rehabilitation (ReCaRe) instrument	Ski et al.	2019
Development and Content Validation of the Pediatric Eating Assessment Tool (Pedi-EAT)	Thoyre et al.	2014
Development and content validity of a screening instrument for gaming addiction in adolescents: The Gaming Addiction Identification Test (GAIT)	Vadlin, S., Åslund, C., & Nilsson, K. W.	2015
Development and Validation of the Just Culture Assessment Tool for Nursing Education	Walker, D., Altmiller, G., Barkell, N., Hromadik, L., & Toothaker, R.	2019
Development and Validation of the Breakthrough Pain Assessment Tool (BAT) in Cancer Patients	Webber, Davies	2014
Development and validation of a learning needs assessment scale: a continuing professional education tool for multiple sclerosis specialist nurses	While, A., Ullman, R., & Forbes, A.	2007

Table 8: Articles on “health app” OR “health games” AND assessment

Title	Author	Year
Naturalistic evaluation of a sport-themed mental health and wellbeing app aimed at men (MindMax), that incorporates applied video games and gamification.	Cheng et al.	2020
What makes a good health “app”? Identifying the strengths and limitations of existing mobile application evaluation tools.	Dawson et al.	2020
Real-World Assessments of mySugr Mobile Health App. Diabetes Technology & Therapeutics	Debong, F., Mayer, H., & Kober, J.	2019
Challenges in Assessing Mobile Health App Quality: A Systematic Review of Prevalent and Innovative Methods	Grundy, Q. H., Wang, Z., & Bero, L. A.	2016
Evaluation of the Effectiveness of Mobile App-Based Stress-Management Program: A Randomized Controlled Trial.	Hwang, W. J., & Jo, H. H.	2019
Suicide prevention and depression apps’ suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines.	Martinengo et al.	2019
Development and testing of a mobile application to support diabetes self-management for people with newly diagnosed type 2 diabetes: a design thinking case study	Petersen, M., & Hempler, N. F.	2017
Mental Health Apps in China: Analysis and Quality Assessment	Shang, J., Wei, S., Jin, J., & Zhang, P.	2019
Availability and quality of mobile health app privacy policies	Sunyaev, A., Dehling, T., Taylor, P. L., & Mandl, K. D.	2015
Understanding the quality, effectiveness and attributes of top-rated smartphone health apps	Wisniewski et al.	2019

5.1 Stage 1: Instrument design

5.1.1 Identification of conceptual framework

Studies vary on the over-all framework applied in developing and validating an instrument for used in health care as shown in Table 9: Conceptual Framework Results. Literature review, qualitative interviews or both and followed by evaluation remain consistent as observed from the literature included in this study.

The framework described by Lynn (1986) which advocated for conducting two stages – development and judgement, has been adapted by all instrument developers. The first stage is where the developer generates the domains and items of what the instrument intends to measure, followed by quantitative judgement of its content based on the assertions of relevance and clarity using a purposive sampling of expert panel. She emphasized that using a two-stage process is fundamental to instrument development in order to determine and quantify content validity.

Table 9: Conceptual Framework Results

Author	Framework
Fredericksen et al.	1. Literature review for item generation 2. Patient concept elicitation interviews 3. Focus group interviews 4. Clinical relevance assessment by the research team and clinicians 5. Cognitive interview for finalization of items 6. Validity testing using existing validated instruments
Halek, M., Holle, D., & Bartholomeyczik, S.	1. Literature review for item generation 2. CVI via expert panel 3. Evaluation study (Feasibility and practicability)
Numminen, O., Katajisto, J., & Leino-Kilpi, H.	1. Literature review for item generation 2. Face validity 3. Content validity using expert panel 4. Pilot testing
Oldland, Botti	1. Focus group interviews 2. Content validity 3. Literature review assessed if framework is current 4. Next study is testing

Polit, D. F., & Beck, C. T.	1. Advocated for the rigorous process of Lynn's method and using the Content Validity Index to validate instruments quantitatively
Pölkki et al.	1. Literature review for item generation 2. Qualitative interview of clinicians 3. Content Validity test 4. Feasibility and clinical utility tests 5. Language translation
Shin, S., & Kang, J.	1. Literature review for item generation 2. Qualitative interviews 3. Content validity test using expert panel (two rounds) 4. Reliability test
Ski et al.	1. Instrument appraisal 2. Face and Content validity using expert panel 3. Test-retest reliability
Thoyre et al.	1. Literature review for item generation 2. Content Validity test for relevance and clarity 3. Cognitive interviews
Vadlin, S., Åslund, C., & Nilsson, K. W.	1. Literature review for item generation 2. Content Validity test using expert panel
Walker et al.	1. Literature review for item generation 2. Content Validity test using expert panel (two rounds) 3. Pilot study
Webber, Davies	1. Literature review for item generation 2. Delphi Process 3. Semi-structured interview of patients 4. Content and construct validity tests 5. Reliability test
While, A., Ullman, R., & Forbes, A.	1. Literature review for item generation and survey of stakeholders 2. Content Validity using expert panel 3. Feasibility test

5.1.2 Item generation

The approach performed on the articles included in the study yielded four distinctive concepts which has been categorized as domains. The concepts of quality of content, technical properties, user-orientation and data security were identified. The quality of content of the app has appeared most across all articles as well as evidence-based information and data security. Based on the issues and characteristics often raised from literatures, items were generated and disseminated into aforementioned domains.

The initial version of the scoresheet consisted of 39 items. Please refer to Table 10: Item Generation Results for the summary.

Table 10: Item Generation Results

Author	Content	User-orientation	Technicality	Security
Cheng et al.	evidence-base recommendation	involving users in development, target user testing		
Dawson et al.	quality of information, scientific evidence, supported by trials, targeted outcomes	user appropriateness, engagement, app cost, 2-way communication between users and developers	functionality, aesthetics, updates	password protection, data protection
Debong, F., Mayer, H., & Kober, J.	evidence-base information, supported by trials	user satisfaction, usability	development should not be one-size fits all	user privacy
Grundy, Q. H., Wang, Z., & Bero, L. A.	evidence-base, testing, supported by trials, guidelines, regulations	involving users in development, target user testing, sociocultural factors of users	methodological development, language support	data privacy & security policy

Hwang, W. J., & Jo, H. H.	Test with target users, effectivity, health information	convenience, usefulness, satisfaction, w-way communication	Usability and feasibility test	protecting user data
Martinengo et al.	adherence to evidence-base, guidelines, trustworthy content, clear health category	target user specification, cost transparency, declaration of affiliation	ability to contact health professionals, updates, language support, two-way communication	password and data protection, data security
Petersen, M., & Hempler, N. F.	need for concrete and simple information, evidence-base	involving users in development, user testing, user support, brief tutorial	technical support, ease of navigation, simple understandable design, ability to integrate with other devices	
Shang, J., Wei, S., Jin, J., & Zhang, P.	specific health information and purpose, health information, based on guidelines	engagement, cultural considerations	accessibility, language support, two-way communication, aesthetics, function	
Sunyaev, A., Dehling, T., Taylor, P. L., & Mandl, K. D.	health information, guidelines, simplicity of statements	informed decisions on data use and cost	technical transparency, developer information	privacy policy

Wisniewski et al.	regulations, guidelines, contents based on research, claim support, evidence-base	user control to their information	technical performance, testing, updates	data privacy & security policy

5.1.3 Structuring the instrument format

A 5-point Likert scale was adopted in the scoresheet and the end points are labelled as 1 (Strongly Disagree) to 5 (Strongly Agree). All item statements were positively worded. This study intends to create an instrument that will measure the degree of how evident the concept of a specific domain item is in the app that is being reviewed. Therefore, using a lesser than 7 or greater than 11-point scale will not compromise the quantitative quality of its measurement nor it will benefit more with a wider variance of responses. Figure 3 shows an excerpt of an item from the scoresheet.

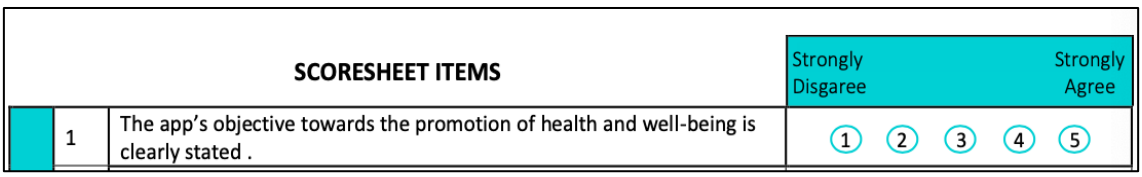

Figure 3: Scoresheet item with five-point Likert scale assessment

To score the items in the scoresheet, reviewers are encouraged to use and inspect the app carefully and rate the items based on their professional objective judgment. This information is clearly stated in the user manual as follows:

- Score 1 (Strongly Disagree) should be given when there is no information provided in relevance to the item.
- Score 5 (Strongly Agree) should be given when the quality of information in relevance to the item is exceptionally described.
- Scores 2 to 4 should be given when information in relevance to the item is present; but lacks completeness and quality.

Reviewers are reminded that some items may refer to laws, directives, guidelines or standards. Thus, it is recommended that reviewers inspect these to be able to provide meaningful scoring and assessment of the item.

The scoresheet uses the percent of maximum possible score method (POMPS) to calculate the domain scores. It provides useful information on the content of the scoresheet represented in Likert scale with the consideration of the lowest and highest possible score (Cohen P, Cohen J, Aiken L & West S 1999). Thus, it is meaningful to the objective and concept of this study. In calculating for POMPS, item scores for each domain across reviewers are added-up and furthered as standardized percentage of the maximum possible score. Please refer to Table 11: POMPS Calculation on Health Content domain with 4 reviewers and the calculation as follows:

Domain score % = total item score – minimum possible score
maximum possible score – minimum possible score

Maximum possible score = 5 x number of domain items x number of reviewers
Minimum possible score = 1 x number of domain items x number of reviewers

Table 11: POMPS Calculation on Health Content domain with 4 reviewers

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Total
Reviewer 1	4	4	4	4	3	4	23
Reviewer 2	4	4	4	5	3	4	24
Reviewer 3	4	4	4	5	3	4	24
Reviewer 4	5	4	5	4	3	4	25
	17	16	17	18	12	16	**96**

Maximum possible score = 5 x 6 x 4 = 120
Minimum possible score = 1 x 6 x 4 = 24
Domain score = 96 – 24 / 120 – 24
= 72 / 96
= 0,75 x 100
= 75%

Domain scores are treated independently. It reflects the score quality of the app based on the specific domain and provide measurable comparison on the app's content and development process. Cut-off scores were not set. Consequently, the app reviewers of scoresheet hold the decision in identifying the level of suitability of the apps being reviewed and if it should be recommended for use to its target users. In any case, an item shall be disregarded from the domain score calculation due to its non-applicability (N/A) as per unanimous decision by the reviewers, the calculation of both maximum and minimum possible scores shall be adjusted accordingly. This decision should be discussed and agreed upon by those involved. An item may be excluded

from the domain score calculation if and only agreed unanimously by the reviewers. The scoresheet also provides reviewers to express their over-all remarks on the app as either they recommend the app if they agree that it showcases good quality content and meets the criteria based on the domains, or otherwise the app needs improvement. In addition, reviewers may also suggest and point-out areas that can be improved on the app.

The user manual was developed for the purpose of guiding users in utilizing the scoresheet. It outlines specific information to help reviewers conduct a systematic assessment of the app as well as identifying if the app being reviewed falls under the medical or non-medical device category prior to proceeding further in using the instrument.

The user manual provides description of the domain items and carefully explains technical terms. It guides users where the information can be usually found either in the app or from the supplementary document provided by the app developer or designers. It also reminds users to check that tests, trials and compliance to specific directives, standards, laws and guidelines were made available with necessary documents for reference. The user manual emphasized that as there are no strict policies nor legislations for non-medical apps, compliance to recommendations can add value to its quality and reliability. It also guides reviewers in any case a specific domain item needs to refer to additional documents for proof and reference purposes. The scoresheet and user manual are to be used hand-in-hand to guide the reviewers in evaluating the app.

5.2 Stage 2: Judgment

5.2.1 Face Validity

The initial version of scoresheet with a total of 39 items was tested by the invited participants (n = 6) on the 16th of March 2020 and rated the instrument based on clarity of instruction, ease-of-use, layout and structure in a 4-point Likert scale. The degree of agreement yielded 83% across the three assertions as shown in Table 12: Face Validity Test Results. The results indicated that the scoresheet format is feasible. Four out of the 6 participants provided feedback as follows:

- “Some of the items in the scoresheet were difficult to evaluate and it helps if information is provided where they can usually be found in the app.”
- “Criteria on security and privacy should be explicitly presented to the users.”
- “Some items can be just added features, maybe they don’t really tell much about the quality of the app but they can be also helpful.”
- “The scoresheet can be shorter.”

Table 12: Face Validity Test Results

Participant	Instruction	Ease-of-use	Layout & Structure
1	3 (Clear)	3 (Easy)	3 (Good)
2	2 (Confusing)	3 (Easy)	3 (Good)
3	4 (Very Clear)	3 (Easy)	3 (Good)
4	3 (Clear)	3 (Easy)	2 (poor)
5	3 (Clear)	2 (Difficult)	4 (Very good)
6	3 (Clear)	3 (Easy)	3 (Good)
Agreement	**83 %**	**83 %**	**83 %**

Based on the test results, the need for user manual has been established to provide guidance to reviewers. After the testing, the scoresheet domain items were discussed with the supervisors and members of the Qvalidi Consortium. Vocabulary and sentence construction were improved and paraphrased in a simpler manner. Some items were re-categorized and moved to Basic Details section for the reason that the quality of the app should not be scored with the presence or absence of the item in question; but rather identified as an added value. For example, the item under technical properties *"The app can be integrated with other applications, IoT devices or other health-related software"* was paraphrased and moved to basic details section for the reason mentioned.

5.2.2 Content Validity

The first round of content validation for the scoresheet and user manual with a total of 34 items commenced on the 6th of June 2020 and concluded on the 8th of October 2020 with 8 expert panel participants. All the participants represented each of the criteria that was set for the purposive sampling. There were 33 items that garnered an I-CVI > 0.78 on relevance except for the item under User-orientation "*The app have attractive graphics and/ or visuals"* (I-CVI = 0.75) which according to literature, based on its I-CVI score should be revised. All items reached content valid scores on clarity of description in the user manual (I-CVI > 0.78). On the scale level, the scoresheet received good universal agreement and average scores (S-CVI/UA = 0,91; S-CVI/Ave = 0,99). Descriptive comments from the experts on vocabulary and additional areas to be discussed on item description in the user manual has been considered to improve the items both in the scoresheet and user manual. Polit, Beck and Owen (2007) observed that the CVI value of 1.00 was acceptable for panels of three or four experts. Whereas, an I-CVI value of greater than 0.78 is suggested to have an excellent content validity due to the consideration of chance agreement for a panel of 5 members or

more. Appendix 4: Content Validation Round 1 Results on Relevance and Appendix 5: Content Validation Round 1 Results on Clarity show the summary of results.

Literature suggest that it is best to perform the second round of content validation by the same expert panel participants as in round one (Lynn M R 1986, Polit & Beck 2006, Halek H 2017). The second round commenced on the 13th of October 2020 and concluded on the 27th of October 2020 with 7 out of 8 experts. Two items were rated on the assertions of relevance and clarity and both items are deemed as content valid with an I-CVI < 0,78 for both assertions of relevance and clarity. Descriptive comments from the experts on item descriptions and use of complex sentences were taken into account to improve the items. Please see Table 13: Content Validation Round 2 Results for Relevance and Table 14: Content Validation Results for Clarity.

Table 13: Content Validation Round 2 Results for Relevance

DOMAIN	ID	E1	E2	E3	E4	E5	E6	E7	Experts in Agreement	I-CVI
Technical Properties	10	1	1	1	1	1	1	1	7	1
User-orientation	15	1	1	1	0	1	1	1	6	0,86

Table 14: Content Validation Round 2 Results for Clarity

DOMAIN	ID	E1	E2	E3	E4	E5	E6	E7	Experts in Agreement	I-CVI
Technical Properties	10	1	1	1	1	1	1	1	7	1
User-orientation	15	1	1	1	1	1	1	1	7	1

The content validated scoresheet and user manual consists of 34 items in total and categorized into Basic Details, Health Content, Technical Properties, User-orientation, Privacy and Safety.

The following sub-sections present the actual contents of the scoresheet and user manual. Formats and lay-out were modified for the purpose of this paper.

5.3 Scoresheet Contents

The Qvalidi tool is used to assess and support the quality of "non-medical" health and well-being apps based on four domains - Health Content, Technical Properties, User-orientation, Privacy & Safety. Please refer to REGULATION (EU) 2017/745 on medical devices if the app has a medical purpose such as monitoring, diagnosing and preventing a disease, injury or handicap.
Please verify if any supplementary documentation has been provided to support claims and compliance.

The Basic Details section pertains to the relevant information that identifies the health app being reviewed, its classification and elements as listed below:

1. App name
2. App version and release date
3. Developer or publisher and publish date
4. Operating platform (iOS, Android, macOS, Windows or others)
5. Cost availability (free, free with in-app purchases, paid)
6. Intended use (health care, educational, leisure time or others)
7. Classification (game, gamified or others)
8. Affiliation (university or university of applied sciences, government, non-governmental organization or others)
9. Funding (university or university of applied sciences, government, non-governmental organization or others)
10. Accessories (e.g.: VR headmount display, controllers, sensors, mobile or medical device, other hardware needed)
11. Integration with other devices, applications or health-related software (e.g.: smart watch, computer, patient information system)

The Health Content section pertains to the assessment of quality and validity of health information in the health app being reviewed. It examines the following:

1. The app's objective towards the promotion of health and well-being is clearly stated.
2. The developers presented background theory or health care concept used in developing the principles of the app.
3. The developers presented how the background theory or health care concept is implemented in the achievement of app's health objective.
4. The developers presented scientific evidence to support its claims on health and well-being.
5. All health information provided in the app is up-to-date in line with the current clinical guidelines.

6. The developers declared that they included or consulted health care experts in the same field as the app is intended for.

The Technical Properties section pertains to the assessment of the technicality, functionality and over-all performance of the health app being reviewed. It examines the following:

7. The app is not only limited to a single operating system (iOS and Android for mobile devices; macOS/ windows for web-based).
8. The developers specified the app's minimum device requirement.
9. The app is available in language/s that is/ are relevant to the intended user group or geographical location (e.g. Available in Finnish, Swedish and English).
10. The app's user interface (UI) effectively directs users towards the desired actions and information.
11. Technical support information is available to the user in the app.
12. The app does not have lengthy advertisements that can potentially disrupt its objective.
13. Relevant tests has been performed and results were presented by the developers to support the technical reliability and usability of the app.

The User-orientation section pertains to the assessment of how the health app and its development process have involved and put consideration to its target users with the following elements:

14. The intended user group of the app is clearly defined (e.g. patient/ non-patient and age group).
15. The app provides means of feedback to and from the user based on user experience.
16. Tutorial, instruction or user guide was provided to the user group.
17. The app is user-friendly to the intended user group.
18. The developer declared the app's compliance with the EU Web Accessibility Directive 2016/2102.

The Privacy and Safety section pertains to the assessment of how the health app and its developers use and protect the user information, privacy and data security with the following elements:

19. Users were warned of any potential risks in using the app (e.g. prolonged use of the app, misused outside the purpose).
20. The app asked for user's consent on collecting data (including cookie policies and personal data).
21. The app provided information on how user data is used, stored and protected.
22. The app stated its compliance with the EU Privacy Code of Conduct on Mobile Health App (please see manual for detailed description).
23. The app stated its compliance with the General Data Protection Regulation (GDPR).
24. The app does not impose any form of ethical risk (religious/ cultural/ sexual in nature).

5.4 User Manual Contents

Introduction

The Qvalidi tool is designed to assess and support the development, evaluation and reporting of health and well-being application programs (apps) that fall under the "non-medical device" category. The tool is comprised of the checklist, scoresheet and user manual.

Please refer to REGULATION (EU) 2017/745 on medical devices if the app has a medical purpose such as monitoring, diagnosing and preventing a disease, injury or handicap. For further information, please see:

- Fimea: Medical Devices https://www.fimea.fi/web/en/medical-devices/legislation
- Finlex® 629/2010. Laki terveydenhuollon laitteista ja tarvikkeista (in Finnish) https://www.finlex.fi/fi/laki/ajantasa/2010/20100629

The term "app" will be used throughout the user manual referring to any application program developed for mobile and web-based platforms.

The user manual will serve as the guide in using the scoresheet and utilizing the Qvalidi tool in general.

Preparation for using Qvalidi tool

App designers and developers can use the Qvalidi Checklist as a structured guide in the development process of their app. The reviewers can use the Qvalidi Scoresheet and User Manual for the evaluation of how the app meets the items identified in the checklist. The scoresheet and user manual should be used hand-in-hand to guide the reviewers in scoring the items.
It is recommended that tests, trials and compliance to specific directives, standards, laws and guidelines are to be supplemented with necessary documents for reference. Chapter 5 - Scoring the Items will guide reviewers in any case a specific domain item needs to refer to additional documents.

Qvalidi recommends having a minimum of two independent reviewers in conducting the assessment.

Who can use Qvalidi tool?

Qvalidi tool is intended to be used by the following:

- **Health care providers** who wish to assess the validity and safety of non-medical apps intended to be used in the promotion of health and well-being to individuals prior to recommendation.

- **Health care and well-being educators** who wish to impart good-practice knowledge through apps.
- **Researchers** who wish to assess the validity and safety of different health and well-being apps under the non-medical device category for research purposes.
- **eHealth and mHealth app developers/ designers** who wish to follow a structured methodology with the use of Qvalidi Checklist and in conducting an internal assessment of their product with the Qvalidi Scoresheet and User Manual.

Content and Structure of Qvalidi Scoresheet

The Qvalidi Scoresheet is comprised of the basic details section and the 24 items which individually represent the four unique domains – Health Content, Technical Properties, User-orientation, Privacy & Safety. In addition, each reviewer is to provide an over-all recommendation of the app and an optional free-text comment or suggestions.

About the Item description section:
This section provides information referring to the specific item. It attempts to elaborate the item further and may provide examples to guide the reviewers on having the best possible understanding of the item's context value.

About the Where to look section:
In this section, reviewers are guided to where the information can be usually found either in the app or from the supplementary document provided by the app developer/ designers. Reviewers may be directed to inspect documents or specific guidelines, directives, laws and standards in order to give the most appropriate score for the item in question. It is of the reviewer's own professional prerogative and level of judgment as he/ she sees fit in scoring the item.

Basic Details
It is recommended that app developers or their representative should fill-out the Basic Details section as shown in figure 4 or otherwise provide the information to the reviewers accurately.

	Basic Details					
i	App Name:			Version:		Release Date:
ii	Developer/ Publisher:					Publish Date:
iii	Operating Platform:	☐ iOS	☐ Android	☐ macOS	☐ Windows	☐ Others (pls. specify):
iv	Cost Availability:	☐ Free	☐ Free with in-app purchase			☐ Paid
v	Intended Use:	☐ Healthcare	☐ Educational	☐ Leisure time		☐ Others (pls. specify):
vi	Classification:	☐ Game	☐ Gamified			☐ Others (pls. specify):
vii	Affiliation: ☐ University/ UAS ☐ Government ☐ NGO ☐ Others Please specify: _____					
viii	Funding: ☐ University/ UAS ☐ Government ☐ NGO ☐ Others Please specify: _____					
ix	Accessories (e.g. VR headmount display, controllers, sensors, medical device, other hardware needed):					
x	Integration with other devices, applications or health-related softwares (e.g. smart watch, computer, patient information system):					

Figure 4: Basic details excerpt from the actual scoresheet

i. App Name, Version and Release Date

Please specify the complete name of the app being reviewed, its latest version and the date the latest version was released.

ii. Developer/ Publisher and Publish Date

Please specify the developer and/ or publisher of the app as well as the date the app was first published for use.

iii. Operating Platform

This section refers to the operating system (OS) of the app. Please mark all that applies in any case the app is available in more than one operating platform (i.e. Available in both iOS and Android, in both Windows and macOS or in all specified choices). If the app is also available in other unspecified operating systems from the Qvalidi scoresheet, “Others” should also be selected, and the operating system of the app should be specified. Some apps may also be available in obscure or lesser known/ used platforms by the general users. It is important that developers also specify these systems. It is possible that some gamified health apps are also

made available in console platforms (i.e. Orbis OS for PlayStation or Windows 10 Core OS for Xbox), they should also be specified.

iv. Cost Availability

This section refers to the availability of the app either for limited or public use. Some apps may only be available in a controlled environment meant for professional use or trained individuals. In both situations, please specify whether the app requires payment, completely free-to-use or free with in-app purchase offers. In any case, the app offers free trial period and requires payment after a certain amount of time, it is considered as requiring payment. The duration of the trial period can be specified should be found necessary.

v. Intended Use

Health care

Some non-medical device apps can be used in supporting health care processes even if they do not function as a medical device. These apps may serve the purpose of providing general information to health care workers and/ or patients, used as a communication platform, provide support or as an assistance tool. However, these apps are not used to diagnose nor mitigate treatment and prevention measures on a disease, injury, medical condition or handicap. Some apps can be used by patients outside health care facilities in support of their care plan. It is important that the app developer declares and emphasizes in any case the use of the app requires a health care professional, a trained practitioner or if the app can be used by patients without any form of supervision. Declarations can be made to the user upon launching the app and through the instruction manual. It is strongly advised that the reviewer confirms these. In both situations, the app is intended for use in health care.

Educational

Apps can be intended for use in schools or other learning facilities to support the learning process of students as its target user group. Gamified health apps are popular amongst young population and schools may utilize these apps to support the educational process of the students. As most apps provide health education in different ways, this option refers to apps that are specifically developed to be utilized in schools or a learning facility as a tool to support its user group in terms of health and well-being.

Leisure time

Apps that are used for leisure time may aim to provide general health or patient education in the promotion of health and well-being practices outside a health care facility or professional supervision (i.e. games that help educate children about good hand hygiene or dangers of

smoking). In some cases, apps can be used in both health care environment and leisure time. For example, apps that are used to assist in relaxation and rest time. It can be used by health care workers, patients and by anyone outside the health care environment (i.e. home).

It is possible that more than one option can be selected to identify the intended use setting of the app being reviewed.

Others

In any case the app is intended to be used in any other settings which are not provided in the options, please mark "Others" and specify the intended use setting.

vi. Classification

Please mark accordingly as per the options provided:

- *Game* if an app engages users to achieve the gameplay objective in a fun and entertaining way. It involves different components such as players, goals, rules, competition and opponents[1,19,20,21]. The goal and rules may aim at challenging users towards specific health and well-being lessons or information by playing.
- *Gamified* when an app applies the elements, mechanics and design of a game such as collecting points or badges, level progression or competitions in a non-game context[5,14]. Gamified apps may aim to encourage or promote motivation towards health and well-being goals.
- In any case the app is not classified as a game nor gamified, please select "Others" and specify the app's classification (i.e. chat app, health dictionary, etc).

vii. Affiliation

Please mark accordingly as per the options provided:

- *University / UAS* – if the app is developed with the university or university of applied sciences (i.e. capstone or thesis projects).
- *Government* – if the app is developed with any form of government agency.
- *NGO (Non-Governmental Organisation)* – if the app is developed with a public organisation (i.e. unions, clubs, private health care institution, etc.).
- *Others* – if the app is developed independently without any affiliation with neither of the provided options (i.e. private or self-funded by the developer).

Names, credential, affiliations of the developers, sponsors, and owners should be specified in the provided field. If the app is developed in cooperation of several stakeholders, choose the organization which has the largest responsibility of the app. If that is impossible to determine, choose "Others" and specify the reason for choosing others.

viii. Funding

Please note that externally funded apps either full or partial, are subject to compliance to applicable guidelines, directives, laws and standards. Specific domain items may require additional documents, please ensure that they are available for reference.

Name/s of funding bodies, grants, sponsors or private funding should be specified in the provided field.

ix. Accessories

Some apps require extra accessories in order to use. Please specify any hardware needed other than mobile or computer in order for the app to achieve its health and well-being objective (i.e. VR (Virtual Reality) apps that need head-mounted display, controller, VR monitor/ projector, smart watch, IoT-integrated health devices, medical devices such oximeter, thermometer, external sensors, etc.

x. Integration

Please specify if the app have optional integration functions with other devices or software (i.e. apple watch, health devices, patient information system used in health institutions, etc.). If the app requires to be integrated or connected to any other form of device in order to be used, please identify details in the Accessories section.

Scoring the Items

In this section, each item in the scoresheet is described. Reviewers are guided what to look for and where it can usually be found. To score the items, it is important to inspect the app carefully and rate the items based on professional objective judgment.

Score 1 (Strongly Disagree) should be given when there is no information provided in relevance to the item.

Score 5 (Strongly Agree) should be given when the quality of information in relevance to the item is exceptionally described and meets all the necessary requirements such as when applied test documentations were presented, compliance to legislation and/ or recommendations.

Scores 2 to 4 should be given when information in relevance to the item is present; but lacks completeness and quality. Some items may refer to laws, directives, guidelines or standards. It is recommended to inspect these to be able to provide suitable scoring and assessment of the item.

Health Content

1. The app's objective towards the promotion of health and well-being is clearly stated.

Item Description:
It is important that developers should be able to describe clearly the primary objective and purpose-of-use of the app. In addition to establishing the health and well-being benefits to its target user group, it determines whether it falls under the medical device category and subject to the Finnish Medical Device Act (629/2010) and other mandatory regulation and directives.

As per the Medical Device Act (629/2010), a medical device is "any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, intended by the manufacturer to be used for human beings for the purpose of: i) diagnosis, prevention, monitoring, treatment or alleviation of disease; ii) diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap; iii) investigation, replacement or modification of the anatomy or of a physiological process; or iv) control of conception."

Please see:

- Fimea: Medical Devices (https://www.fimea.fi/web/en/medical-devices)
- Finlex® 629/2010. Laki terveydenhuollon laitteista ja tarvikkeista (in Finnish, https://www.finlex.fi/fi/laki/ajantasa/2010/20100629)

Where to look:
Should be found in the description or upon launching the application program. The objective can be further described with a supplementary document.

2. The developers presented background theory or health care concept used in developing the principles of the app.

Item Description:
This item refers to background theory on health or well-being utilized in the modelling of the app and its operating principles. It is possible that more than one theory has been used to develop the principles of the apps.

Where to look:
Usually found in the description or supplementary document.

3. The developers presented how the background theory or health care concept is implemented in the achievement of app's health objective.

Item Description:
The developers should be able to describe how the implemented background theory supports the achievement of the health or well-being objective of the app.

Where to look:
Usually found in the description or supplementary document.

4. The developers presented scientific evidence to support its claims on health and well-being.

Item Description:
The developers should be able to present research evidence on the health and well-being impacts of the app to the user. The key results and data source should also be presented. If evidence is not available of the current app, the developers should state what is the research basis where the app is leaning. For example, the app is aiming to increase communication between the employer and employee, and it is a knowledge based on evidence that better communication between these two is increasing the job satisfaction. Hence, the app itself has not been studied if it does increase job satisfaction. However, the developer may state that based on evidence (with references) that better communication is increasing job satisfaction and that the app is aiming to increase communication.

Where to look:
Research evidence, tests and results may be mentioned in the description section of the application program. However, supplementary documents or publication weblinks should be provided by the developers for reference. Otherwise, it cannot be considered as fully existent in the app nor valid.

5. All health information provided in the app is up to date in line with the current clinical guidelines.

Item Description:
Any health information provided in the app, directly stated or implied should be supported by either theoretical literature or scientific research. Health information may also be based or adopted from current clinical guidelines or evidence-based information.

It is important that the assessor is aware of the current clinical guidelines pertaining to the health information included in the app. A well-founded app should be able to provide references or citations.

It is recommended to inspect the app rigorously if in any case the app contain other health and well-being information (i.e. application uses information on the health effects of tobacco, but this information is not presented to the user in the form of a statement).

Where to look:
Reference and citations may be available in the description section of the application program or supplementary documents may also be provided by the developers.

6. The developers declared that they included or consulted health care experts in the same field as the app is intended for .

Item Description:
The involvement of health care sector expertise provides additional health content value to the app. This may be in a form of actual participation in the development, testing or consultation. Names of experts, professionals, agencies or organisations in some cases, are included in the information provided within the app.

Where to look:
Usually found in the description or supplementary document.

Technical Properties

7. The app is not limited to a single operating system only (iOS and Android for mobile devices; macOS/ windows for web-based).

Item Description:
It is recommended that the availability of the app should not be limited to one operating system only depending on the device.

Where to look:
Usually found in the description or in the app download site information.

8. The developers specified the app's minimum device requirement.

Item Description:
This item refers to the compatibility of the app with the physical device (i.e. mobile or computer). It should include the operating system (OS) versions, file size of the app, required free disk space in the device and processor (computer/ web-based apps).

Where to look:
This information can be found in the app's download site and in app's information section.

9. The app is available in language/s that is/ are relevant to the intended user group or geographical location (e.g. Available in Finnish, Swedish and English).

Item Description:
It is important that an app is available in languages based on the geographical location of the target group. In Finland, apps available in Finnish, Swedish and English meet this criterion excellently.

Where to look:
Should be indicated in the app's information section or upon launching the app. It is usually in the form of language selection feature.

10. The app's user interface (UI) effectively directs users towards the desired actions and information.

Item Description:
The app's general aesthetics greatly influences the user's interaction with the app. User Interface (UI) such as screen appearance, colour schemes, layout, looks and feel of the graphics affect user interaction. A good UI should be intuitive enough to direct users effectively to desired actions and information – navigation buttons, icons and other graphical elements as well as the interactive behaviour (i.e. swiping left or right, scrolling up or down) that most users are familiar with[4,16,17].

Some game and gamified elements such as interactive behaviours may not be familiar or explicitly visible to users. Developers should be able to introduce them in an effective and easy way such as providing in-app tutorials upon first-time use of the app or predictive text notifications.

Where to look:
Reviewers are advised to use the app long enough to acquire the actual hands-on user experience of the app being reviewed. As some elements can be subjective, reviewers may consider in assessing the app in the perspective of the target user group.

11. Technical support information is available to the user in the app.

Item Description:
The app should be able to provide information for users to contact the app developers for any technical issue or inquiry that may occur upon its use.

Where to look:
Usually provided as a link in the app. It may be in a form-type or the developer's email address is displayed.

12. The app does not have lengthy advertisements that can potentially disrupt its objective.

Item Description:
Apps that are fee-to-use may contain advertisements that may interrupt the use of its content and features for a certain amount of time, usually in several seconds or upon clicking the exit ("X") button. Depending on the main objective the app, the assessor should weigh upon whether the timeframe to resume in using the app disrupts its quality.

The type of advertisements should also be inspected whether it is appropriate to the user group of the app. Some apps may contain advertisement that may impose harm or unhealthy information (e.g. advertisements that are not appropriate for specific age group).

Where to look:
It usually pops-up whilst using the app.

13. Relevant tests has been performed and results were presented by the developers to support the technical reliability and usability of the app.

Item Description:
It is important that developers declared all tests (usability, reliability, cloud, automated or manual testing) and/ or trials performed during the development process of the app, including where and when the tests were performed and with which user group. Test or trial results should also be presented by the developers for reference.

Where to look:
Tests and/ or trials can be mentioned in the description of the app and should be supported by supplementary documents for reference.

User-orientation

14. The intended user group of the app is clearly defined (i.e. patient/ non-patient and age group).

Item Description:
The intended user group (by population, age or health condition) should be well described in the app. In any case the use of the app involves mediators such as health care professionals or educators, it should be made clear upon prior to further use of the app.

Where to look:
Information on the app's target user group should be available immediately on the description of the app prior to downloading and upon launching the app. It can also be described further in any supplementary documentation when available.

15. The app provides means of feedback to and from the user based on user experience.

Item Description:
This item refers to the capability of the app to provide and support feedback to and from the user. Apps may provide feedback to the user on its content, features and functions upon use (e.g. sounds, notifications, etc.). The app should also be able to provide opportunity for users to send feedback, review, comments or suggestions based on their over-all user experience such as the app's interaction, ease-of-use, enjoyment or fulfilling user expectations about the app.

Where to look:
Sounds and notification feedback can be determined upon use of the app. The user feedback section is usually provided as a link in the app. It may be in a form-type or the developer's email address is displayed.

16. Tutorial, instruction or user guide was provided to the user group.

Item Description:
The app should be able to provide clear instructions on how to effectively use the app in order to meet the health and well-being objectives. Recommendations such as screen-time, ideal frequency of use can also be described. Tutorials can also be made available to guide users in achieving optimal results of what the app is

intended for. The user guide provides complete information about the app, its purpose, objective and instructions on its use.

Where to look:
Some apps provide in-app tutorials upon first-time use. Separate user guide or manual may also be available.

17. The app is user-friendly to the intended user group.

Item Description:
The ease-of-use of the app should be assessed based on the perspective of the target user group. Language, vocabulary, technical features, appropriateness of tasks (whenever applicable), aesthetics (layout, visual appeal) contribute to over-all user experience.

Where to look:
Results of usability tests performed can be considered upon assessment of the item. It is encouraged however, to inspect the app carefully whilst the assessor uses the app themselves and evaluate the ease-of-use of the app with the utmost consideration of the users' condition and ability (e.g. cognitive functions, mental, physical ability).

18. The developer declared the app's compliance with EU Web Accessibility Directive 2016/2102.

Item Description:
Apps funded by public authority either partial or full is subject to obligatory compliance to the EU Web Accessibility Directive 2016/2102. The Regional Administration of Southern Finland advises, directs and supervises the implementation of the directive.

Apps that do not fall under this compliance, however; can abide by the directive to demonstrate good quality. This means that the app does not discriminate the accessibility to for example visually impaired, colour-blind or limited mobility users.

Where to look:
Reviewers are encouraged to be familiar with the *EU Web Accessibility Directive 2016/2102.* Apps may be customisable such as the option to change font size, change colour schemes to suit the user's needs. It is often found in app settings or options.

Please see:

- Finlex® 306/2019. Laki digitaalisten palvelujen tarjoamisesta. https://www.finlex.fi/fi/laki/alkup/2019/20190306
- EU 2016/2102. Euroopan parlamentin ja neuvoston direktiivi (EU) 2016/2102. Available at: https://eur-lex.europa.eu/legal-content/FI/TXT/HTML/?uri=CELEX:32016L2102&from=FI
- W3C Recommendation 2011. Authorized Translation: Web Content Accessibility Guidelines (WCAG 2.0). Available at: https://www.w3.org/Translations/WCAG20-fi/

Privacy & Safety

19. Users were warned of any potential risks in using the app (i.e. prolonged use of the app, misused outside the purpose).

Item Description:
Potential risks associated with using the app may occur outside its intended purpose. It is important that developers are able to identify the associated risks and provide encouragement to users in avoiding such risks.

Some health information is susceptible to misconception and misinterpretation of the user when unsupervised. The app developers should be able to provide encouragement to consult for professional help in any case users feel the need of requiring further information or uncertainties.

Where to look:
Usually found in the description, disclosure or as a warning in the app. It may also be available in instructions.

20. The app asked for user's consent on collecting data (including cookie policies, personal data).

Item Description:
A cookie is a file sent from a website and stored in a user's device browser via an internet connection. It can track the user's activity and can even show advertisements. Users should be able to agree or disagree on the collection of their data and storing of cookie trackers in their device where they have installed the app.

Where to look:
A cookie disclaimer should be displayed upon launching or using the app. Users should be prompted whether they accept the cookie policy or not.

21. The app provided information on how user data is used, stored and protected.

Item Description:
The app should be able to provide detailed information on how user's information is collected, processed and protected from any form of breached.

Where to look:
Usually found in a disclaimer or terms and conditions section of the app.

22. The app stated its compliance with the EU Privacy Code of Conduct on Mobile Health App.

Item Description:
Seeking for the approval from the European Data Protection Board on the EU Privacy Code of Conduct on Mobile Health App is voluntary to app developers. Nevertheless, it increases the app's "trusted" value in the promotion of health and well-being and provides advantage. The code covers guidance and protection principles on the following:

1. User consent
2. Purpose limitation and data minimisation
3. Privacy by design and by default
4. Data subject rights and information requirements
5. Data retention
6. Security measures
7. Advertisement
8. Use of personal data and secondary purposes
9. Data disclosure to third parties
10. Data transfers
11. Data breach
12. Data gathered for children

For more detailed information on EU Privacy Code of Conduct on Mobile Health App, please see: https://ec.europa.eu/digital-single-market/en/privacy-code-conduct-mobile-health-apps

Where to look:
Usually included on the "Terms and Conditions" of using the app.

23. The app stated its compliance with the General Data Protection Regulation (GDPR).

Item Description:
The General Data Protection Regulation (GDPR) applies to all apps that collect and process personal data in EU. It aims to protect the user's Personally Identifiable Information (PII). Generally, the GDPR regulates the following:

- Consent on access, storage, processing and deletion of user's personal data.
- Privacy and security protocols of user data.
- Transparency of data protection breach and risks.

For more detailed information on GDPR, please see:

- https://eur-lex.europa.eu/eli/reg/2016/679/oj
- https://www.finlex.fi/fi/laki/ajantasa/2018/20181050 (in Finnish)

Where to look:
Usually included on the "Terms and Conditions" of using the app. Users may be asked to agree by clicking a checkbox or a button to proceed with the sign-in process.

24. The app does not impose any form of ethical risk (religious/ cultural/ sexual in nature).

Item Description:
Ethical considerations should be made clear in the app in any case the content may possibly impose any form of ethical risk. Proper precautions on sensitive health issues and conditions should be conveyed by the app, as well as special attention to vulnerable groups (i.e. children and adolescents). Some apps may have consulted an ethics committee during the development and/or research process.

Where to look:
Usually found as a disclaimer and/ or in the terms and conditions section of the app.

Over-all Recommendation

Reviewers may provide their over-all recommendation on the app based on the following options:

Yes, I recommend the app.

If the reviewer agrees that the app being reviewed showcases good quality content and meets the criteria based on Qvalidi domains, therefore recommends the app for use.

The app needs improvement (reviewer may specify areas that can be improved on the provided field below).

If the reviewer objectively sees the need for improving the app prior to use by its target user group. Reviewers may give suggestions in the provided field.

Reviewers are encouraged to assess the app objectively based on their professional judgement. Domain scores serve the purpose of providing measurable comparison on the app's content and development process.

Members of the Qvalidi Consortium 2021

Lotta Hamari, PT, PhD	Postdoctoral Researcher Department of Nursing Science University of Turku
Anni Pakarinen, RN, PhD	Senior Researcher Development Manager Postdoctoral Researcher Department of Nursing Science University of Turku
Heidi Parisod, PHN/ RN, PhD	Postdoctoral Researcher Department of Nursing Science University of Turku
Minna Aromaa, MD	Docent/ Adjunct Professor Department of Public Health University of Turku
Sanna Salanterä, RN, PhD	Professor, Vice Dean Department of Nursing Science University of Turku
Ville Leppänen	Professor of Software Engineering Department of Future Technologies University of Turku
Magda Skogberg, PHN/ RN, MHSc	Lead Trainer Oy Apotti Ab

6 Discussion

In this study, the development of the scoresheet and user manual to quantitatively assess the degree of quality of health apps were described. It was not surprising to know that at present, there are no quality assessment tools that can be used to measure the quality of non-medical device health apps. Hence, the purpose of this undertaking.

Although it can be seen that a clear gap has been apparent when it comes to European-wide standards such as having regulations or policies that can control the presence of these apps in the market, or for public use back in 2014 during the publication of the Green Paper on mHealth (EU Commission 2014) and again in 2017 when the Report of the Working Group on mHealth Assessment (EU Commission 2017) was released, nothing was concluded to consensus in order to formulate any form regulating policy.

The need for systematic examination of the constantly growing mHealth devices and technology has been determined since then. To date, amidst the Covid-19 pandemic, the prevalence of digitized medical and health care practices has been largely increasing. Medical devices are regulated; but its counterpart category – the non-medical devices are left to multiply without a validation or examination of its content quality and health promoting claims. Policies and regulations are meant to exist not to hinder growth and advancement, but rather to ensure that safe and good quality innovations are being catered for all.

The literature review performed on health apps yielded results that examine the important qualities that should be paid attention to. Although the studies are scarce during the time of this endeavor, it has been identified that evidence-based content and adherence to health care guidelines are two of the most important domains that health apps should cover. Otherwise, there will be no grounds for validity nor reliability for these devices. Additionally, involvement of target users in the development and trials, technicalities of the app and its adherence to data privacy and security should be examined and ensured.

Although literature varies when it comes to rules in developing assessment instruments for used in health care context, it is common to most previous studies to initially direct instrument developers to refer to evidence and empirical information to generate its contents, with the utmost consideration of its purpose and scope. As Lynn mentioned (1986), it has two stages – the development and judgement.

The items in the developed scoresheet of this study were based on the most discussed qualities of health apps that health care professionals and providers often examine as written in scientific papers. Additionally, applicable policies and regulations were also inspected in order to ensure that items are relevant and parallel

to any existing guidelines and recommendations in the European Union, even though there are no policies that regulates the health apps under the non-medical device category.

The development of the scoresheet has considered the reduction of response burden and designed with straightforward and simple statements. Additionally, the user manual contributes value to the practicality of using the scoresheet. Users will not need additional training nor video materials in order to use the instrument.

The user manual serves as a guide for further clarification and elaboration of items. It provides examples and direction to where users can locate what a particular item in the scoresheet specifically refers to. It also includes references to applicable policies and directives in any case an app falls into a regulated category or the app's purpose warrants for supplemental documents such as declaration of tests performed to a certain target group or any proof of conducted procedure to support their claims.

As a result of the judgement phase, the scoresheet and user manual yielded a total of 34 items categorized into basic details, health content, technical properties, user-orientation, privacy and safety. Due to Covid-19 restriction measures imposed in the country, physical interviews were difficult to arrange for the author. The rigorous search of validated instruments with similar purpose in EU as with this study, yielded no results. Thus, there were no means to do a comparison study of measurement instruments. On the other hand, descriptive feedbacks during the face validity testing were collected to gather inputs from a small sample of intended users of the scoresheet.

According to expert's judgement, measuring the quality of apps based on aesthetics or how attractive the graphics are, is not of utmost relevance when they are primarily intended to be used in health care context. The author has put some thoughts and gathered that as in any health app, good aesthetics is a given prerequisite in its own design and development process. Therefore, if we are to measure the degree of quality of a health app based on its aesthetics, it will not provide a conclusive measurement of quality.

In any study that involves the opinions of experts, subjectivity and bias pose a threat to validity. However, it can be claimed and noted that the ratings given by experienced experts from the fields are based on objective judgement. Moreover, purposive and systematic expert panel criteria determine the level of objectivity and quality of results yielded from this stage (Rubio et al., 2003).

The Content Validity Index (CVI) has been used in this study instead of the other well-known measure to quantify content validity, the Lawshe's (1975) Content Validity Ratio (CVR). Due to the unexplainable criteria of CVR minimum values against having 8 or 9 experts as shown in Figure 5: Lawshe's CVR Value per Number of Panelist, this was not considered to measure content validity in this study. On the other hand,

the CVI has provided objective results on whether to retain, revise or omit an item. CVI is also easier to calculate and understand compared to CVR.

PERSONNEL PSYCHOLOGY

TABLE 1
Minimum Values of CVR and CVR$_t$
One Tailed Test, p = .05

No. of Panelists	Min. Value*
5	.99
6	.99
7	.99
8	.75
9	.78
10	.62
11	.59
12	.56
13	.54
14	.51
15	.49
20	.42
25	.37
30	.33
35	.31
40	.29

Figure 5: Lawshe's CVR Value per Number of Panelist (1975)

7 Conclusion

In conclusion, developing an instrument requires an extensive process and rigorous evaluation. The author has developed a scoresheet and user manual that can be used to measure the degree of quality of health apps. The study was carried out adopting a scientific methodology as described by scientific literature and has displayed rigor in its development and evaluation stages. It has established face and content validity for all domain items achieving good CVI scores for both item and scale levels that resulted from the two expert panel sessions. The purposive sampling criteria used for the expert panel showcased a relevant and careful inclusion of expertise in the domains of health and technology, thus contributes significantly on the objective judgement during the content validation of the domain items in the scoresheet and user manual.

Face and content validity are important forms of validation that should be performed most especially if the assessment instrument is the first of its kind in its area of implementation. These forms of validity are prerequisites for developing an assessment instrument for use in health care context. They are very important for clinicians and researchers who require high-quality measurement instrument (Polit, Beck and Owen, 2007). It adds further rigor to the development and evaluation process of the instruments. However, reliability, feasibility and acceptability studies are still needed prior to full implementation. Due to time constraints and the current COVID-19 restrictions in Finland during the time of this study, it was challenging for the author to carry out the testing procedures with proper sampling methodology. Hence, the limitation of this study.

The measurement for test-retest reliability should be performed in the next study period in order to establish the consistency and accuracy of the instruments' performance, that the results remain unchanged on two different time points from the same sample group (Vilagut G, 2014). The correlations will provide the reliability estimate of the developed instruments in addition to statistical analyses of internal consistency and coefficients (Bartone P, 2007). Moreover, an adequate stakeholder sampling of atleast 5 participants per item should be adopted (Hair et al., 1998).

The feasibility and acceptability of the developed scoresheet and user manual can be explored from the perspectives of the defined stakeholders, using mixed-methods approach. Data collection can be performed through semi-structured interviews and structured questionnaires designed for the context of exploring the purpose of the study. To capture the satisfaction and acceptability of the instrument, interviews following a theme frame can be carried out, audio recorded and transcribed in verbatim. Qualitative data can be analysed using directed content analyses (Hsieh & Shannon, 2005), whilst the quantitative data can be analysed using descriptive statistics and correlation coefficients analysis (Gray, Grove & Sutherland, 2017)

The intended scoresheet and user manual has the potential to provide health care professionals, educators, app and game developers and designers an instrument that can adequately evaluate health apps quantitatively for its degree of quality and validity.

Consents from all the participants in both the face validity testing and expert panel were acquired. Please see Appendix 4: Letters to seek consent for study participation. All members of the expert panel declared that their participation is of their own personal accords and that there were no needs to acquire permission from their respective affiliations. Permission to conduct study from the Qvalidi Consortium and declaration of Intellectual Property Rights Agreement from the University of Turku, Department of Nursing Science were also acquired.

References

Abdollahpour, E., Nejat, S., Nourozian, M. & Majdzadeh, R. (2010). The process of content validity in instrument development. Iranian Epidemiology. 2010;6(4):66–74

Albrecht, U., Von Jan, U., & Pramann, O. (2013). Standard reporting for medical apps. Studies in Health Technology and Informatics, 190, 201–203. https://doi.org/10.3233/978-1-61499-276-9-201

Albrecht, U., Hillebrand, U., & Von Jan, U. (2018). Relevance of trust marks and CE labels in German-language store descriptions of health apps: Analysis. JMIR mHealth and uHealth, 6(4), e10394–e10394. https://doi.org/10.2196/10394

Bartone, P. (2007). Test-retest reliability of the dispositional resilience scale-15, a brief hardiness scale. Psychological Reports, 101(3), 943–944. https://doi.org/10.2466/PRO.101.3.943-944

Bishop, P. A., & Herron, R. L. (2015). Use and Misuse of the Likert Item Responses and Other Ordinal Measures. International journal of exercise science, 8(3), 297–302.

Boateng, G., Neilands, T., Frongillo, E., Melgar-Quinonez, H., & Young, S. (2018). Best Practices for Developing and Validating Scales for Health,Social, and Behavioral Research:A Primer. Frontiers in Public Health, 6, 149–149. https://doi.org/10.3389/fpubh.2018.00149

Brown-Johnson, C., Berrean, B., & Cataldo, J. (2015). Development and usability evaluation of the mHealth Tool for Lung Cancer (mHealth TLC): A virtual world health game for lung cancer patients. Patient Education and Counseling, 98(4), 506–511. https://doi.org/10.1016/j.pec.2014.12.006

Carmines, E., & Zeller, R. (1979). Reliability and validity assessment. SAGE.

Cheng, V., Davenport, T., Johnson, D., Vella, K., Mitchell, J., & Hickie, I. (2020). Naturalistic evaluation of a sport-themed mental health and wellbeing app aimed at men (MindMax), that incorporates applied video games and gamification. Internet Interventions : the Application of Information Technology in Mental and Behavioural Health, 20, 100306–100306. https://doi.org/10.1016/j.invent.2020.100306

Cohen, P., Cohen, J., Aiken, L., & West, S. (1999). The Problem of Units and the Circumstance for POMP. Multivariate Behavioral Research, 34(3), 315–346. https://doi.org/10.1207/S15327906MBR3403_2

Connell, J., Carlton, J., Grundy, A., Taylor Buck, E., Keetharuth, A. D., Ricketts, T., Barkham, M., Robotham, D., Rose, D., & Brazier, J. (2018). The importance of content and face validity in instrument development: lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation, 27(7), 1893–1902. https://doi.org/f10.1007/s11136-018-1847-y

Costa Alexandre, O. (2011). Content validity in the development and adaptation processes of measurement instruments. Ciência & saude coletiva, 16(7), 3061–3068. https://doi.org/10.1590/S1413-81232011000800006

Davis, L. (1992). Instrument review: Getting the most from a panel of experts. Applied Nursing Research, 5(4), 194–197. https://doi.org/10.1016/s0897-1897(05)80008-4

Dawson, R. M., Felder, T. M., Donevant, S. B., McDonnell, K. K., Card, E. B. 3rd, King, C. C., & Heiney, S. P. (2020). What makes a good health “app”? Identifying the strengths and limitations of existing mobile application evaluation tools. Nursing Inquiry, 27(2), e12333. https://doi.org/10.1111/nin.12333

Debong, F., Mayer, H., & Kober, J. (2019). Real-World Assessments of mySugr Mobile Health App. Diabetes Technology & Therapeutics, 21(S2), S235–S240. https://doi.org/10.1089/dia.2019.0019

EU 2016/679 Regulation of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation). https://eur-lex.europa.eu/eli/reg/2016/679/oj

EU 2017. Report of the Working Group on mHealth Assessment Guidelines. February 2016 - March 2017. Euroopan Unioni, Bryssel.

EU Council Directive 93/42/EEC of 14 June 1993 concerning medical devices OJ L 169 of 12 July 1993. Available at: https://ec.europa.eu/growth/single-market/european-standards/harmonised-standards/medical-devices_en

European Commission on CE Marking. (2020). Internal Market, Industry, Entrepreneurship and SMEs. Available at: https://ec.europa.eu/growth/single-market/ce-marking_en

European Commission. (2014). Green Paper on mobile health ("mhealth"). https://ec.europa.eu/digital-single-market/en/news/green-paper-mobile-health-mhealth

European Commission. (2018). Privacy Code of Conduct on mobile health apps. https://ec.europa.eu/digital-single-market/en/privacy-code-conduct-mobile-health-apps

European Commission. (2017). Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and93/42/EEC (Text with EEA relevance.) https://eur-lex.europa.eu/legal-content/ EN/ALL/?uri=CELEX:32017R0745

Finlex® 5.12.2018/1050 Tietosuojalaki 1050/2018. https://www.flnlex.fi/fl/laki/ajantasa/2018/20181050

Fredericksen, R. J., Fitzsimmons, E., Gibbons, L. E., Dougherty, S., Loo, S., Shurbaji, S., Batey, D. S., Avendano-Soto, S., Mathews, W. C., Christopoulos, K., Gray, J. R., Grove, S. K., & Sutherland, S. (2017). *The practice of nursing research: Appraisal, synthesis and generation of evidence*. Elsevier.

Grundy, Q. H., Wang, Z., & Bero, L. A. (2016). Challenges in Assessing Mobile Health App Quality: A Systematic Review of Prevalent and Innovative Methods. American Journal of Preventive Medicine, 51(6), 1051–1059. https://doi.org/10.1016/j.amepre.2016.07.009

Halek, H. (2017). Development and evaluation of the content validity, practicability and feasibility of the Innovative dementia-oriented Assessment system for challenging behaviour in residents with dementia. BMC Health Services Research, 17(1), 554–554. https://doi.org/10.1186/s12913-017-2469-8

Hair Jr., J.F., Anderson, R. E., Tatham, R.L. & Black, W.C. (1998). Multivariate data analysis 5th Edition Prentice Hall Upper Saddle River

Hamari, L., Parisod, H., Pakarinen, A., Skogberg, M., Aromaa, M., Leppänen, V. & Salantera, S. (2020). Digitaalisten terveys- ja hyvinvointisovellusten kehittäminen, arviointi ja raportointi: Qvalidi 2019-tarkistuslistan kehittäminen ja sisältö. Hoitotiede 2020, 32 (1), 52-66.

Haynes, S. N. et al. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. Psychological assessment. [Online] 7 (3), 238–247.

Henson. P., David, G., Albright, K. & Torous, J. (2019). Deriving a practical framework for the evaluation of health apps. The Lancet Digital Health 1(2), e52-e54. (11)

Holden, R.R. (2010). Face validity. Corsini Encyclopedia of Psychology.

Hsieh, H. F. & Shannon, S. E. (2005). Three approaches to qualitative content analysis. Qualitative Health Research, 15(9), 1277-1288.

Hwang, W. J., & Jo, H. H. (2019). Evaluation of the Effectiveness of Mobile App-Based Stress-Management Program: A Randomized Controlled Trial. International Journal of Environmental Research and Public Health, 16(21). https://doi.org/10.3390/ijerph16214270

Jake-Schoffman D.E., Silfee VJ., Waring M.E., Boudreaux E.D., Sadasivam R.S., Mullen S.P, Carey J.L., Hayes R.B., Ding E.Y., Bennett G.G. & Pagoto S.L. (2017). Methods for evaluating the content, usability, and efflcacy of commercial mobile health apps. JMIR mHealth and uHealth 5(12), e190. (12)

Johns, R. (2010). Likert Items and Scales. Survey Question Bank: Method Fact Sheet 1. University of Strathclyde, Glasgow United Kingdom.

Kunter, B. (2006). Who is the expert?: Construct and criteria validity of student and teacher ratings of instruction. Learning Environments Research, 9(3), 231–251. https://doi.org/10.1007/s10984-006-9015-7

Lawshe, C. (1975). A Quantitative Approach To Content Validity. Personnel Psychology, 28(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Linturi, R. & Kuusi, O. (2018). Suomen sata uutta mahdollisuutta 2018-2037. Yhteiskunnan toimintamallit uudistava radikaali teknologia. Eduskunnan tulevaisuusvaliokunnan julkaisu 1/2018. Tulevaisuusvaliokunta, Helsinki

Lynn, M.R. (1986). Determination and quantification of content validity. Nurs Res. 1986;35:382–5.

Martinengo, L., Van Galen, L., Lum, E., Kowalski, M., Subramaniam, M., & Car, J. (2019). Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. BMC Medicine, 17(1), 1–12. https://doi.org/10.1186/s12916-019-1461-z

Mayer, K. H., Mugavero, M. J., Crane, P. K., & Crane, H. M. (2019). Development and content validation of the Multifactoral assessment of perceived social support (MAPSS), a brief, patient-reported measure of social support for use in HIV care. AIDS Care, 31, 1–9. https://doi.org/10.1080/09540121.2019.1612005

Ministry of Justice, Finland. (2003). Language Act 423/2003. https://www.finlex.fi/en/laki/kaannokset/2003/en20030423.pdf

Moors, K. (2014). The Effect of Labeling and Numbering of Response Scales on the Likelihood of Response Bias. Sociological Methodology, 44(1), 369–399. https://doi.org/10.1177/0081175013516114

Murphy, K. (2005). Psychological testing: principles and applications (International ed.). Prentice Hall.

Numminen, O., Katajisto, J., & Leino-Kilpi, H. (2019). Development and validation of Nurses' Moral Courage Scale. Nursing Ethics, 26(7–8), 2438–2455. https://doi.org/10.1177/0969733018791325

Nunnally, J.C., Bernstein, I.H. (1994). Psychometric theory. 3rd ed. New York: McGraw-Hill.

Oldland, E., Botti, M., Hutchinson, A., & Redley, B. (2020). A framework of nurses' responsibilities for quality healthcare — Exploration of content validity. Collegian (Royal College of Nursing, Australia), 27(2), 150–163. https://doi.org/10.1016/j.colegn.2019.07.007

Pakarinen, A., Parisod, H., Linden, I., Aromaa, M., Smed, J., Leppanen, V., & Salantera, S. (2017). Usability of a gamified application to promote family wellbeing in child health clinics. In P. Tuomi & A. Perttula (Eds.), Proceedings of the 1st International GamiFIN Conference, Pori, Finland, May 9-10, 2017. CEUR Workshop Proceedings 1857, 141- 149.

Pakarinen, A., Flemmich, M., Parisod, H., Selänne, L., Hamari, L., Aromaa, M., Leppänen, V., Löyttyniemi, E., Smed, J., & Salanterä, S. (2018). Protocol for digital intervention for effective health promotion of small children—A cluster randomized trial. Journal of Advanced Nursing, 74(7), 1685–1699. https://doi.org/10.1111/jan.13561

Petersen, M., & Hempler, N. F. (2017). Development and testing of a mobile application to support diabetes self-management for people with newly diagnosed type 2 diabetes: a design thinking case study. BMC Medical Informatics & Decision Making, 17, 1–10. https://doi.org/10.1186/s12911-017-0493-6

Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. Research in Nursing & Health, 29(5), 489–497. https://doi.org/10.1002/nur.20147

Polit, D. F., Beck, C. T. & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Research in Nursing & Health, 30, 459 – 467.

Pölkki, T., Korhonen, A., Axelin, A., Saarela, T., & Laukkala, H. (2014). Development and preliminary validation of the Neonatal Infant Acute Pain Assessment Scale (NIAPAS). International Journal of Nursing Studies, 51(12), 1585–1594. https://doi.org/10.1016/j.ijnurstu.2014.04.001

Rubio, D., Berg-Weger, M., Tebb, S., Lee, E., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. Social Work Research, 27(2), 94–104. https://doi.org/10.1093/swr/27.2.94

Shang, J., Wei, S., Jin, J., & Zhang, P. (2019). Mental Health Apps in China: Analysis and Quality Assessment. JMIR MHealth and UHealth, 7(11), e13236. https://doi.org/10.2196/13236

Shin, S., & Kang, J. (2019). Development and Validation of a Person-Centered Perioperative Nursing Scale. Asian Nursing Research, 13(3), 221–227. https://doi.org/10.1016/j.anr.2019.07.002

Ski, C. F., Jones, M., Astley, C., Neubeck, L., Thompson, D. R., Gallagher, R., Du, H., & Clark, R. A. (2019). Development, piloting and validation of the Recommending Cardiac Rehabilitation (ReCaRe) instrument. Heart & Lung, 48(5), 405–413. https://doi.org/https://doi.org/10.1016/j.hrtlng.2019.04.008

STM. (2016). Digitalisaatio terveyden ja hyvinvoinnin tukena. Sosiaali- ja terveysministeriön digitalisaatiolinjaukset 2025. Sosiaali- ja terveysministeriön julkaisuja 2016.

Stoyanov, S.R., Hides, L., Kavanagh, D.J., Zelenko, O., Tjondronegoro, D. & Mani, M. (2015). Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth and uHealth 3(1), e27.

Sunyaev, A., Dehling, T., Taylor, P. L., & Mandl, K. D. (2015). Availability and quality of mobile health app privacy policies. Journal of the American Medical Informatics Association, 22(e1), e28-33. https://doi.org/10.1136/amiajnl-2013-002605

Thoyre, S. M., Pados, B. F., Park, J., Estrem, H., Hodges, E. A., McComish, C., Van Riper, M., & Murdoch, K. (2014). Development and Content Validation of the Pediatric Eating Assessment Tool (Pedi-EAT). American Journal of Speech-Language Pathology, 23(1), 46–59. https://doi.org/10.1044/1058-0360(2013/12-0069)

Tomlinson, M., Rotheram-Borus, M., Swartz, L., & Tsai, A. (2013). Scaling Up mHealth: Where Is the Evidence? PLoS Medicine, 10(2), e1001382–e1001382. https://doi.org/10.1371/journal.pmed.1001382

Tsang, S., Royse, C., & Terkawi, A. (2017). Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi Journal of Anaesthesia, 11(5), 80–S89. https://doi.org/10.4103/sja.sja_203_17

Vadlin, S., Åslund, C., & Nilsson, K. W. (2015). Development and content validity of a screening instrument for gaming addiction in adolescents: The Gaming Addiction Identification Test (GAIT). Scandinavian Journal of Psychology, 56(4), 458–466. https://doi.org/10.1111/sjop.12196

Valvira. (2019). Lääkinnällisten laitteiden asetukset. https://www.valvira.fi/terveydenhuolto/terveysteknologia/laakinnallisten-laitteiden-asetukset

Vilagut, G. (2014). Test-Retest Reliability. In: Michalos A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_3001

W3C Recommendation. (2011). Authorized Translation: Web Content Accessibility Guidelines (WCAG) 2.0 https://www.w3.org/Translations/WCAG20-fi/

Walker, D., Altmiller, G., Barkell, N., Hromadik, L., & Toothaker, R. (2019). Development and Validation of the Just Culture Assessment Tool for Nursing Education. Nurse Educator, 44(5), 261–264. https://doi.org/10.1097/NNE.0000000000000705

Webber, K., Davies, A., Zeppetella, G., & Cowie, M. (2014). Development and Validation of the Breakthrough Pain Assessment Tool (BAT) in Cancer Patients. Journal of Pain and Symptom Management, 48(4), 619–631. https://doi.org/10.1016/j.jpainsymman.2013.10.026

While, A., Ullman, R., & Forbes, A. (2007). Development and validation of a learning needs assessment scale: a continuing professional education tool for multiple sclerosis specialist nurses. Journal of Clinical Nursing, 16(6), 1099–1108. https://doi.org/10.1111/j.1365-2702.2007.01693.x

Wisniewski, H., Liu, G., Henson, P., Vaidyam, A., Hajratalli, N. K., Onnela, J.-P., & Torous, J. (2019). Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. Evidence-Based Mental Health, 22(1), 4–9. https://doi.org/10.1136/ebmental-2018-300069

Wynd, C., Schmidt, B., & Schaefer, M. (2003). Two Quantitative Approaches for Estimating Content Validity. Western Journal of Nursing Research, 25(5), 508–518. https://doi.org/10.1177/0193945903252998

Yasini, M., Beranger, J., Desmarais, P., Perez, L., & Marchand, G. (2016). mHealth Quality: A Process to Seal the Qualified Mobile Health Apps. Studies in Health Technology and Informatics, 228, 205–209.

Zamanzadeh, G. (2015). Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. Journal of Caring Sciences, 4(2), 165–178. https://doi.org/10.15171/jcs.2015.017

Zhang, M., Yu, J., Shen, W., Zhang, Y., Xiang, Y., Zhang, X., Lin, Z., & Yan, T. (2020). A mobile app implementing the international classification of functioning, disability and health rehabilitation set. BMC Medical Informatics and Decision Making, 20(1), 12–12. https://doi.org/10.1186/s12911-020-1019-1

Appendices

Appendix 1: Content Validation Round 1 Results on Relevance

DOMAIN	ID	SME1	SME2	SME3	SME4	SME5	SME6	SME7	SME8	Experts in Agreement	I-CVI	UA
Basic Details	1	1	1	1	1	1	1	1	1	8	1	1
	2	1	1	1	1	1	1	1	1	8	1	1
	3	1	1	1	1	1	1	1	1	8	1	1
	4	1	1	1	1	1	1	1	1	8	1	1
	5	1	1	1	1	1	1	1	1	8	1	1
	6	1	1	1	1	1	1	1	1	8	1	1
	7	1	1	1	1	1	1	1	1	8	1	1
	8	1	1	1	1	1	1	1	1	8	1	1
	9	1	1	1	1	1	1	1	1	8	1	1
	10	1	1	1	1	1	1	1	1	8	1	1
Health Content	11	1	1	1	1	1	1	1	1	8	1	1
	12	1	1	1	1	1	1	1	1	8	1	1
	13	1	1	1	1	1	1	1	1	8	1	1
	14	1	1	1	1	1	1	1	1	8	1	1
	15	1	1	1	1	1	1	1	1	8	1	1
	16	1	1	1	1	1	1	1	1	8	1	1
Technical Properties	17	1	0	1	1	1	1	1	1	7	0,88	0
	18	1	1	1	1	1	1	1	1	8	1	1
	19	1	1	1	1	1	1	1	1	8	1	1
	20	1	1	1	1	1	1	1	1	8	1	1
	21	1	1	1	1	1	1	1	1	8	1	1
	22	1	1	1	1	1	1	1	1	8	1	1
	23	1	1	1	1	1	1	1	1	8	1	1
User-orientation	24	1	1	1	1	1	1	1	1	8	1	1
	25	1	1	1	1	1	1	1	1	8	1	1
	26	1	0	1	1	1	1	1	1	7	0,88	0
	27	1	1	1	1	1	1	1	1	8	1	1
	28	1	0	1	1	0	1	1	1	6	0,75	0
Privacy & Safety	29	1	1	1	1	1	1	1	1	8	1	1
	30	1	1	1	1	1	1	1	1	8	1	1
	31	1	1	1	1	1	1	1	1	8	1	1
	32	1	1	1	1	1	1	1	1	8	1	1
	33	1	1	1	1	1	1	1	1	8	1	1
	34	1	1	1	1	1	1	1	1	8	1	1
												31
Proportion of Relevance		1,00	0,91	1,00	1,00	0,97	1,00	1,00	1,00	S-CVI/Ave	0,99	
		Average proportion of relevant items across 8 experts 0,99								S-CVI/UA		0,91

Appendix 2: Content Validation Round 1 Results on Clarity

DOMAIN	ID	SME1	SME2	SME3	SME4	SME5	SME6	SME7	SME8	Experts in Agreement	I-CVI
Basic Details	1	1	1	1	1	1	1	1	1	8	1
	2	1	1	0	1	1	1	1	1	7	0,88
	3	1	1	0	1	1	1	1	1	7	0,88
	4	1	1	0	1	1	1	1	1	7	0,88
	5	1	1	1	1	1	1	1	1	8	1
	6	1	1	0	1	1	1	1	1	7	0,88
	7	1	1	1	1	1	1	1	1	8	1
	8	1	1	1	1	1	1	1	1	8	1
	9	1	1	1	1	1	1	1	1	8	1
	10	1	1	1	1	1	1	1	1	8	1
Health Content	11	1	1	1	1	1	1	1	1	8	1
	12	1	0	1	1	1	1	1	1	7	0,88
	13	1	0	1	1	1	1	1	1	7	0,88
	14	1	1	1	1	1	1	1	1	8	1
	15	1	1	1	1	1	1	1	1	8	1
	16	1	1	1	1	1	1	1	1	8	1
Technical Properties	17	1	1	1	1	1	1	1	1	8	1
	18	1	1	1	1	1	1	1	1	8	1
	19	1	1	1	1	1	1	1	1	8	1
	20	1	0	1	1	1	1	1	1	7	0,88
	21	1	1	1	1	1	1	1	1	8	1
	22	1	1	1	1	1	1	1	1	8	1
	23	1	1	0	1	1	1	1	1	7	0,88
User-orientation	24	1	1	1	1	1	1	1	1	8	1
	25	1	0	1	1	1	1	1	1	7	0,88
	26	1	1	1	1	1	1	1	1	8	1
	27	1	1	1	1	1	1	1	1	8	1
	28	1	1	1	1	1	1	1	1	8	1
Privacy & Safety	29	1	1	1	1	1	1	1	1	8	1
	30	1	1	1	1	1	1	1	1	8	1
	31	1	1	1	1	1	1	1	1	8	1
	32	1	1	1	1	1	1	1	1	8	1
	33	1	1	1	1	1	1	1	1	8	1
	34	1	1	1	1	1	1	1	1	8	1

Appendix 3: Sources used for user manual

1. Adams, E. (2014). Fundamentals of game design (3rd ed.). Berkeley, CA: New Riders.
2. Cooper, A. et al. About Face: The Essentials of Interaction Design, John Wiley & Sons, Incorporated, 2014. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/kutu/detail.action?docID=1762072.
3. Deterding, S., Khaled, R., Nacke L.E. & Dixon D. Gamification: Toward a definition. InCHI 2011 gamification workshop proceedings 2011 May 7 (Vol. 12). Vancouver BC, Canada.
4. Grissom, R. J., & Kim, J. J. (2011). Effect sizes for research: Univariate and multivariate applications, second edition. Retrieved from https://ebookcentral.proquest.com
5. Encyclopedia of Information Science and Technology, Fourth Edition. (2017). Information Science Reference (Isr).
6. Johnson, Daniel, & Janet Wiles. "Effective Affective User Interface Design in Games." Ergonomics 46.13-14 (2010): 1332–1345. Web.
7. Knijnenburg, Bart P. et al. "Explaining the User Experience of Recommender Systems." User modeling and user-adapted interaction 22.4-5 (2012): 441–504. Web.
8. NordicEdu Oy. (2015). FUME (Version 1.2.1) [Mobile application software]. https://itunes.apple.com/
9. Prensky, M. (2002). The motivation of gameplay: The real twenty-first century learning revolution. On the Horizon, 10(1), 5–11. https://doi.org/10.1108/10748120210431349
10. Smed, J. & Hakonen, H. (2003) *Towards a definition of a computer game.* Turku: Turku Centre for Computer Science.
11. Suits, B. (1967). WHAT IS A GAME? Philosophy of Science, 34(2), 148–156. https://doi.org/10.1086/288138

Appendix 4: Letters to seek consent from study participants

Face Validity Testing Participants via Google Forms

Dear participants,

The Qvalidi tool is designed for mobile and web-based apps that do not fall under the ¨Medical Device¨ category. Please refer to REGULATION (EU) 2017/745 if the app has a medical purpose such as monitoring, diagnosing and preventing a disease, injury or handicap.

Declaration of Informed Consent to participate in the testing of Qvalidi Scoresheet

You are invited to participate in the testing of Qvalidi Scoresheet, given that you will exercise objective assessment in the role of the intended user of the tool. This is a part of a student's master's thesis in MDP Future Health and Technology, being conducted at the University of Turku.

Participation
Your participation in this activity is voluntary and anonymous under the course of Knowledge Management and Health Technology Supporting Clinical Work (HOIT6701).

Benefits
There will be no direct tangible compensation in participating in this activity. Your participation will contribute to the development of the scoresheet design for the above-mentioned project.

Risks
There are no foreseeable risks in participating in this activity.

Confidentiality
Information gathered will be stored and managed in a password protected spreadsheet. Only statistical data will be used and presented in the study. No personal identity will be used nor disclosed at any stage, only that you are a student of this study program in the University of Turku.

You are giving your informed consent to participate by proceeding to the assessment procedure.

Expert Panel Participants via webropol

Welcome to the Content Validation of Qvalidi Tool 2020 – Scoresheet and User Manual

Dear experts,

You are about to participate in the content validation of the Qvalidi Scoresheet items. This is a part of a student's master's thesis in MDP Future Health and Technology, coordinated with the University of Turku and Qvalidi Consortium.

Rest assured, that your anonymity will be preserved at all stages. Information gathered via this questionnaire will be stored and managed in a password protected spreadsheet. Only statistical data will be used and presented in the study and no personal identity will be used nor disclosed. The results may be made available to industry and/or others in the future or the results may be needed for the purposes of publishing, research or teaching. This study adheres to Yliopistolaki 558/2009 (www.finlex.fi/fi/laki/smur/2009/20090558) and Tietosuojalaki (www.finlex.fi/fi/laki/ajantasa/2018/20181050#L2P4).

Each participant has been provided with the Qvalidi Scoresheet and User Manual in electronic format for reference. This webropol questionnaire serves the purpose of validating the items in the scoresheet and divided into two parts as follows:

Basic details section
The task is to rate each item based on the assertions of *relevance* and *clarity as per the user manual* in a four-point Likert-type scale. Please refer to the user manual provided in rating the clarity of the items. Experts are also provided with the option to give comment or suggestions in the questionnaire as found necessary.

Domain items
The task is to rate each item based on the assertions of *relevance* and *clarity as described in the user manual*, in a four-point Likert-type scale. Please refer to the user manual provided in rating the clarity of the items. Experts are also provided with the option to give comment or suggestions in the questionnaire as found necessary.

The results for each round will be processed and presented in the succeeding round.

If you consent to the above-mentioned statement, please click the AGREE button below.

Appendix 5: IPR Agreement from University of Turku

AGREEMENT ON NON-EXCLUSIVE RIGHTS TO USE THE RESULTS

This agreement ("Agreement") has been concluded between the following parties:

1. Turun yliopisto, (0245896-3) a university organized and existing under the laws of Finland, having its principal place of business at Yliopistonmäki, FI-20014 TURKU, FINLAND ("University"), and
2. Undersigned, who have taken part in the development of Qvalidi Tool during 2020 –2023 ("Project").

The Project is organized and coordinated by the department of Nursing Science. The purpose of this Agreement is to grant rights on the Results created as a part of the Project from the Undersigned to the University. The University may make the Results available to industry and/or others in the future or the Results may be needed for the purposes of publishing, research or teaching, and thus granting of the rights is required.

Therefore, the University and the Undersigned agree as follows:

1. "Results" means the application and any related pieces of work in whatever form, and all intellectual property rights related to those that are created by the Undersigned in connection with the Project.

2. The Undersigned have the copyright on the Results created by them. The Undersigned grant to the University right to use the Results in the scope and for the purposes described in the preamble and in section 3.

3. The Undersigned grant to the University non-exclusively the right to use the economic rights of the Results. The said rights include but are not limited to right to reproduce, make available to the public, revise, transform, translate, build upon, copy, publish, redistribute, exploit and transfer and right to sublicense the Results to third parties. The Undersigned shall retain the copyright on the Result and shall have the right to use the Results for whatever purposes as they see fit.

4. The Undersigned shall be given appropriate credits according to good practice when the University publishes the Results.

5. The Undersigned shall not be paid any compensation for the transfer of the rights described in this Agreement.

6. The transferred right is perpetual, and it may not be cancelled.

7. The laws of Finland shall govern the construction and enforcement of this Agreement. In the event of controversy, claim or dispute arising out of or relating to any provision of this Agreement, the Undersigned and University shall try to settle those conflicts amicably between themselves. Should the parties fail to settle the matter, the dispute shall be finally and exclusively settled by the District court of Varsinais-Suomi.

8. Done in duplicate, one original for Turun yliopisto and the other original for the Undersigned.

9. This Agreement supersedes all prior agreements and understandings (whether written or oral) between the parties of this Agreement, or any of them, with respect to the subject matter hereof.

On behalf of Turun yliopisto

Place and date _____

_____

Helena leino-Kilpi
Head of the Department of Nursing Science
University of Turku

On behalf of the Undersigned

Place and date _____

_____

Sanna Salanterä
Person in Charge
Department of Nursing Science
University of Turku

_____

Kaile Kubota
Student of MDP in Future Health & Technology
Department of Nursing Science
University of Turku

Appendix 6: Content Validated Scoresheet

The QValiDi tool is used to assess and support the quality of "non-medical" health and well-being apps based on four domains - Health Content, Technical Properties, User-orientation, Privacy & Safety. Please refer to REGULATION (EU) 2017/745 on medical devices if the app has a medical purpose such as monitoring, diagnosing and preventing a disease, injury or handicap.

Please verify if any supplementary documentation has been provided to support claims and compliance.

Basic Details		
App Name:	Version:	Release Date:
Developer/ Publisher:		Publish Date:
Operating Platform: ☐ iOS ☐ Android ☐ macOS ☐ Windows ☐ Others (pls. specify):		
Cost Availability: ☐ Free ☐ Free with in-app purchase ☐ Paid		
Intended Use: ☐ Healthcare ☐ Educational ☐ Leisure time ☐ Others (pls. specify):		
Classification: ☐ Game ☐ Gamified ☐ Others (pls. specify):		
Affiliation: ☐ University/ UAS ☐ Government ☐ NGO ☐ Others Please specify: _____		
Funding: ☐ University/ UAS ☐ Government ☐ NGO ☐ Others Please specify: _____		
Accessories (e.g. VR headmount display, controllers, sensors, mobile or medical device, other hardware needed):		
Integration with other devices, applications or health-related softwares (e.g. smart watch, computer, patient information system):		
How to rate: **Score 1 (Strongly Disagree)** should be given when there is no information provided in relevance to the item. **Score 5 (Strongly Agree)** should be given when the quality of information in relevance to the item is exceptionally described. **Scores 2 to 4** should be given when information in relevance to the item is present; but lacks completeness and quality.		

		SCORESHEET ITEMS	Strongly Disagree				Strongly Agree
I. Health Content	1	The app's objective towards the promotion of health and well-being is clearly stated .	1	2	3	4	5
	2	The developers presented background theory or healthcare concept used in developing the principles of the app.	1	2	3	4	5
	3	The developers presented how the background theory or healthcare concept is implemented in the achievement of app's health objective.	1	2	3	4	5
	4	The developers presented scientific evidence to support its claims on health and well-being.	1	2	3	4	5
	5	All health information provided in the app is up-to-date inline with the current clinical guidelines .	1	2	3	4	5
	6	The developers declared that they included or consulted healthcare experts in the same field as the app is intended for .	1	2	3	4	5
II. Technical Properties	7	The app is not only limited to a single operating system (iOS and Android for mobile devices; macOS/ windows for web-based).	1	2	3	4	5
	8	The developers specified the app's minimum device requirement.	1	2	3	4	5
	9	The app is available in language/s that is/ are relevant to the intended user group or geographical location (eg: Available in Finnish, Swedish and English).	1	2	3	4	5
	10	The app's user interface (UI) effectively directs users towards the desired actions and information.	1	2	3	4	5
	11	Technical support information is available to the user in the app.	1	2	3	4	5
	12	The app does not have lengthy advertisements that can potentially disrupt its objective.	1	2	3	4	5
	13	Relevant tests has been performed and results were presented by the developers to support the technical reliability and usability of the app.	1	2	3	4	5
III. User -Orientation	14	The intended user group of the app is clearly defined (eg: patient/ non-patient and age group).	1	2	3	4	5
	15	The app provides means of feedback to and from the user based on user experience .	1	2	3	4	5
	16	Tutorial, instruction or user guide was provided to the user group.	1	2	3	4	5
	17	The app is user-friendly to the intended user group.	1	2	3	4	5
	18	The developer declared the app's compliance with the EU Web Accessibility Directive 2016/2102.	1	2	3	4	5
IV. Privacy & Safety	19	Users were warned of any potential risks in using the app (e.g. prolonged use of the app, misused outside the purpose).	1	2	3	4	5
	20	The app asked for user's consent on collecting data (including cookie policies and personal data).	1	2	3	4	5
	21	The app provided information on how user data is used, stored and protected.	1	2	3	4	5
	22	The app stated its compliance with the EU Privacy Code of Conduct on Mobile Health App(please see manual for detailed description).	1	2	3	4	5
	23	The app stated its compliance with the General Data Protection Regulation (GDPR) (please see manual).	1	2	3	4	5
	24	The app does not impose any form of ethical risk (religious/ cultural/ sexual in nature).	1	2	3	4	5

Acknowledgement

It is of my genuine pleasure to express my utmost gratitude to my University of Turku supervisors Anni Pakarinen and Lotta Hamari for imparting their knowledge, expertise and encouragement that I was able to accomplish this endeavour. You serve as an inspiration for me to yearn for greater things that can bring valuable outcome to many. Thank you for believing in me.

To Prof. Sanna Salanterä, for the knowledge you have guided me to acquire through the challenges during our research seminars. I wouldn't wish to have it in any other way. Also, to Prof. Zhang Yuxia of Fudan University for the guidance and to the rest of the defence committee.

To my classmates from Fudan University and University of Turku, thank you for the fruitful two years of fun and learning.

To my better half, thank you for the support and understanding.