

Portuguese Translation and validation of the user rating scale for mobile applications in the health area (uMARS)

Tradução para o português e validação da escala de avaliação do usuário para aplicativos móveis na área da saúde (uMARS)

Traducción al portugués y validación de la escala de calificación de usuarios para aplicaciones móviles en el área de la salud (uMARS)

Received: 05/25/2023 | Revised: 06/01/2023 | Accepted: 06/05/2023 | Published: 06/10/2023

Silvia Regina Gralha

ORCID: <https://orcid.org/0000-0003-2432-8054>
Universidade Federal de Ciências da Saúde de Porto Alegre, Brazil
E-mail: silvia.gralha@ufcsa.edu.br

Otávio Neves da Silva Bittencourt

ORCID: <https://orcid.org/0000-0002-4638-560X>
Universidade Federal de Ciências da Saúde de Porto Alegre, Brazil
E-mail: otavion@ufcsa.edu.br

Abstract

New health apps can improve patients' lives. However, the rapid proliferation of technological solutions often does not allow a correct assessment of the quality of these tools. For these technologies, it is necessary to carry out more studies and evaluation instruments that stimulate additional research and practices to evaluate the effect of currently available tools and guarantee the quality and usability of mobile technologies. The User Mobile App Rating Scale (uMARS) is a tool that provides a comprehensive, multidimensional assessment of app quality. However, this tool is not available in Portuguese. This study aims to culturally adapt and validate the rating scale questionnaire for users from English to Portuguese. First, a qualitative study was carried out in the translation and validation of the questionnaire. In a second step, the questionnaire was made available to users of the Meu Clínicas© application. Quantitative analysis was performed using Cronbach's Alpha coefficient reliability measure. Results: The 26 questions of the questionnaire were translated in full following a methodology, as well as, all translated items were revised for grammatical terms, verb tenses and synonyms. Content validity was performed through the analysis of five specialists. The study sample consisted of 5 key users and administrative professionals from the Hospital. The instrument was applied from October to November 2022. The Portuguese version of the uMARS had a total Cronbach's alpha of 0.78. The Brazilian version of this scale demonstrated semantic, idiomatic, conceptual and cultural equivalence with the original English version, as well as adequate validity and reliability.

Keywords: Translation; Mobile applications; Health; Assessment instrument.

Resumo

Novos aplicativos de saúde podem melhorar a vida dos pacientes. No entanto, a rápida proliferação de soluções tecnológicas muitas vezes não permite uma avaliação correta da qualidade dessas ferramentas. Para essas tecnologias, é necessário realizar mais estudos e instrumentos de avaliação que estimulem pesquisas e práticas adicionais para avaliar o efeito das ferramentas atualmente disponíveis e garantir a qualidade e usabilidade das tecnologias móveis. O User Mobile App Rating Scale (uMARS) é uma ferramenta que fornece uma avaliação abrangente e multidimensional da qualidade do aplicativo. Este estudo tem como objetivo adaptar culturalmente e validar o questionário da escala de classificação para usuários do inglês para o português. Primeiramente, foi realizado um estudo qualitativo na tradução e validação do questionário. Em uma segunda etapa, o questionário foi disponibilizado aos usuários do aplicativo Meu Clínicas©. A análise quantitativa foi realizada por meio da medida de confiabilidade do coeficiente Alpha de Cronbach. As 26 questões do questionário foram traduzidas na íntegra seguindo uma metodologia, bem como, todos os itens traduzidos foram revisados quanto a termos gramaticais, tempos verbais e sinônimos. A validade de conteúdo foi realizada por meio da análise de cinco especialistas. O instrumento foi aplicado no período de outubro a novembro de 2022. A versão adaptada para a língua portuguesa do uMARS apresentou alfa de Cronbach total de 0,78. A versão brasileira dessa escala demonstrou equivalência semântica, idiomática, conceitual e cultural com a versão original em inglês, bem como adequada validade e confiabilidade.

Palavras-chave: Tradução; Aplicativos móveis; Saúde; Instrumento de avaliação.

Resumen

Las nuevas aplicaciones de salud pueden mejorar la vida de los pacientes. Sin embargo, la rápida proliferación de soluciones tecnológicas muchas veces no permite una correcta evaluación de la calidad de estas herramientas. Para estas tecnologías, es necesario realizar más estudios e instrumentos de evaluación que estimulen investigaciones y prácticas adicionales para evaluar el efecto de las herramientas actualmente disponibles y garantizar la calidad y usabilidad de las tecnologías móviles. La Escala de calificación de aplicaciones móviles del usuario (uMARS) es una herramienta que proporciona una evaluación integral y multidimensional de la calidad de la aplicación. Este estudio tiene como objetivo adaptar culturalmente y validar el cuestionario de escala de calificación para usuarios de inglés a portugués. En primer lugar, se realizó un estudio cualitativo en la traducción y validación del cuestionario. En un segundo paso, el cuestionario se puso a disposición de los usuarios de la aplicación Meu Clínicas®. El análisis cuantitativo se realizó utilizando la medida de confiabilidad del coeficiente Alfa de Cronbach. Las 26 preguntas del cuestionario fueron traducidas en su totalidad siguiendo una metodología, así como todos los ítems traducidos fueron revisados en cuanto a términos gramaticales, tiempos verbales y sinónimos. La validez de contenido se realizó a través del análisis de cinco especialistas. El instrumento se aplicó de octubre a noviembre de 2022. La versión portuguesa del uMARS tuvo un alfa de Cronbach total de 0,78. La versión brasileña de esta escala demostró equivalencia semántica, idiomática, conceptual y cultural con la versión original en inglés, además de adecuada validez y confiabilidad.

Palabras clave: Traducción; Aplicaciones móviles; Salud; Instrumento de evaluación.

1. Introduction

The literature on mHealth has proliferated over the last couple of years and, until now, has mainly focused on the effectiveness and efficiency of the usability of healthcare apps (Melin et al., 2020). According to LeBeau et al. (2019), applications called mHealth (mobile health) have beneficial potential in various fields of health, for some reasons such as low cost for the user, accessibility at any time, and convenience for the user do not require assistance from healthcare professionals to operate successfully and can allow the user to customize their experience. With several health applications available worldwide, it is of paramount importance to evaluate their quality, due to their potential influence on health decisions and outcomes (Bardus et al., 2020).

For Messner et al. (2020), star ratings and app store descriptions often provide insufficient, or even false, information about app quality, although they are popular with end users. Zhou et al. (2019), state that, among the methods of evaluating the usability of mobile apps, the usability questionnaire is the most used due to its simplicity of execution and data analysis. In practice, authors choose to use well-validated usability questionnaires designed for general software systems or create their usability questionnaire according to general usability assessment guidelines.

Because of the importance of the subject, some authors have proposed evaluation methodologies, a scenario in which the Mobile Application Rating Scale (MARS) evaluation model stands out, which is a scale widely used to evaluate the quality and content of applications in the health area (Davalbhakta et al, 2020; Bardus et al., 2020; Domnich et al., 2016). MARS is a multidimensional instrument, indicated to be applied by specialists, to assess the quality of applications in the health area (Stoyanov et al., 2015).

One year after the creation of MARS, the same authors created a new version aimed at users of the application, called uMARS, in which the “u” means user in Portuguese (Stoyanov et al., 2015). In Brazil, there are many surveys of software evaluation in the health area. However, no national studies of adaptation and validation of the uMARS questionnaire were found.

Therefore, the previous statements lead to a question: does uMARS, an evaluation instrument recognized in the scientific community, when translated and adapted into Portuguese, have acceptable reliability for evaluating applications in the health area in Brazil? The present research project aims to present a translated version in Portuguese (Brazilian) of uMARS, which is a known, valid and reliable tool to evaluate health-related applications available in the app stores.

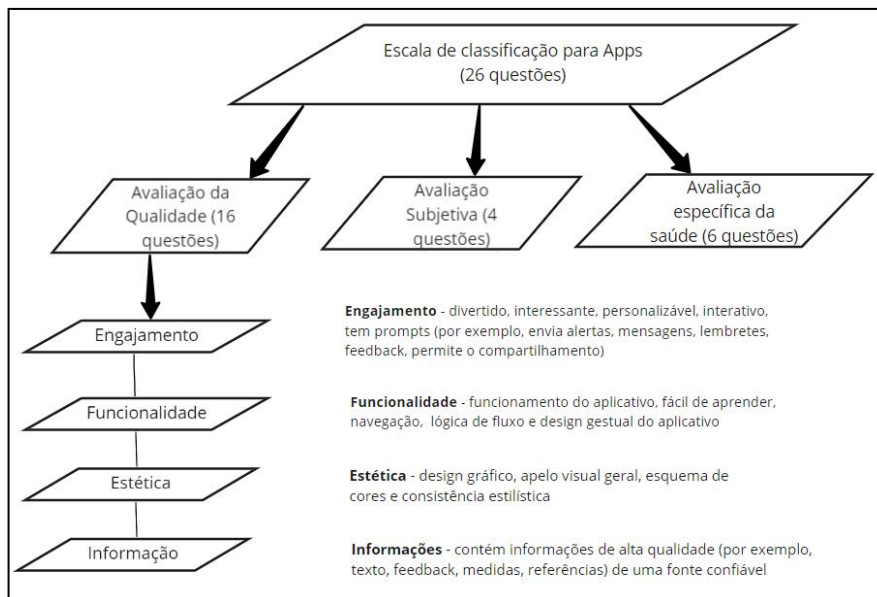
2. Methodology

This research consists of subjecting the study objects to the influence of certain variables under controlled conditions known by the investigator to observe the results that the variable produces in the object (GIL, 2002). The first stage of this research has an exploratory-qualitative character, as it aims to understand and translate the uMARS assessment tool. The purpose is to deepen its translation and meaning to contribute to building a better understanding of the issues to be applied to users of mobile health applications. The second stage is a study of a quantitative nature, as it requires statistical treatment. Quantitative research is characterized by the use of quantification, both in terms of information collection and in their treatment through statistical techniques (Richardson, 1999).

The Mobile App Rating Scale (MARS) “is the only tool currently available that provides a comprehensive and multidimensional assessment of app quality” (Bardus et al., 2020). The original questionnaire in English (Stoyanov et al., 2016) has already been translated and validated into Italian (Domnich et al., 2016), Spanish (Payo et al., 2019), German (Messner et al., 2020) and Arabic (Bardus et al., 2020). However, this tool is not available in Portuguese. To use it, training and experience in mHealth and in the relevant health area are required (Stoyanov et al., 2016). In this way, the author created a simpler version than MARS, with the objective of serving the end user, called uMARS.

According to Stoyanov et al. (2016), uMARS provides 20 items included in 5 subscales in total, 4 objective quality subscales - engagement, functionality, aesthetics, and information quality - and 1 subjective quality subscale that has 6 items and is added to measure the impact perceived by users of the evaluated application. In Figure 1, the structure of the questionnaire is presented.

Figure 1 - uMARS questionnaire structure.



Source: Elaborated by the authors (2022).

The uMARS provides the valuable ability to gather information from users about mobile applications. The quality of mobile applications during the development and testing process can result in general improvements in their quality (Stoyanov et al., 2016).

The uMARS instrument uses the Likert scale, which associates numerical values to expressions of qualitative appreciation such as "little favorable" to "very favorable" (Vagias, 2006). Among the existing scales to measure attitudes, one of the most used in research is the Likert scale (Bermudes et al., 2016). Although the Likert scale used a five-point scale, other

response variations are adequate, including the exclusion of the neutral response (Clason, 1994), as is the case with the questionnaire used in this research.

2.1 The Meu Clínicas© App

The Meu Clínicas© application, developed by the Information and Communication Technology (ICT) team at the Hospital de Clínicas de Porto Alegre (HCPA), aims to qualify the assistance service, facilitate access to useful information by patients and improve the time of care professionals (Menegotto et al., 2019).

Currently, the platform provides information such as laboratory test results, scheduled appointments, care process documents, and the virtualization of the Health Unic System (SUS) card, allowing easy access to the hospital's outpatient clinic.

Self-scheduling appointments online was a great innovation for SUS patients. The Meu Clínicas© patient application can be downloaded from mobile app stores (Google Play or Apple App Store) and accessed via the internet.

2.2 Translation and Cross-Cultural Adaptation

Beaton et al. (2000) developed recommendations for the translation and adaptation of instruments these are currently widely used and were adopted in this study. According to the authors, the translation and cultural adaptation of instruments must be carried out in six stages:

Stage I - Translation: In this step, two bilingual translators, both with Brazilian Portuguese as their mother tongue, must produce two translated versions (T1 and T2) independently.

Stage II - Synthesis: The translators of the TV version of the instrument and the researcher participate. In this step, the synthesis of the results of the T1 and T2 translations is performed, producing a common T1-2 translation.

Stage III - Back-translation: In this stage, two bilingual translators, fluent in English and Brazilian Portuguese, participate. The instrument will be translated (version T1-2) back into the original language, independently and completely blind to the original version and the concepts of the construct, generating two back-translated versions (RT1 and RT2).

Stage IV - Expert Committee: The expert committee identifies the content validity by assessing the semantic, idiomatic, cultural, and conceptual equivalence of each item in the questionnaire.

Stage V - Pre-Test: In order to verify the understanding of the instrument by the target population and verify its applicability, a pre-test is carried out.

Stage VI - Submission and appreciation of all written reports to the creators of the instrument: In this stage, all reports and forms are sent to the authors of the instrument to verify that the recommended stages were followed and the reports reflected positively in the process.

2.3 Evaluation of Validity and Reliability

Souza, Alexandre, and Guirardello (2017) state that “reliability is the ability to reproduce a result consistently, in time and space. While validity refers to the property of an instrument to measure exactly what it is intended to do.” Validity is a concept with less evident theoretical development than the concept of reliability. The Content Validity Index (CVI) is a method widely used in the health area (Waltz, 1981) and measures the proportion or percentage of experts who agree on certain aspects of the instrument and its items (Alexandre, 2011).

The analysis of the reliability of the Portuguese version of the uMARS was performed by analyzing the internal consistency of its items, using Cronbach's alpha coefficient. The alpha is a correlation coefficient whose premise is that the items that make up the instrument are positively related to each other. Therefore, they must measure the same attribute. For the

calculation, the answers were assigned weights equivalent to a Likert-type scale with five points, in which 1 means “I strongly disagree” and 5 “I strongly agree”.

3. Results and Discussion

The original uMARS instrument was sent to two different translation companies. While the companies worked on the translations, invitations to specialists were sent out. The professional profiles of the invited specialists were: systems analysts, developers, administrators, and project leaders.

In total, three rounds of evaluations were required to obtain agreement among experts for content validation and semantic evaluation. In the end, the questionnaire was sent for validation of the translated questionnaires. All responses are shown in Table 1, where TD = Totally Disagree, PD = Partially Disagree, N = Neither Disagree nor Agree, PA = Partially Agree, and TA = Totally Agree.

Table 1 - Experts questionnaire result.

Questions	TD	PD	N	PA	TA	CVI
Did the questions have the same meaning or are they highly similar in both languages?				5		100%
Does the language of the translated questions present a level of difficulty comparable to the words of the original item?		1			4	80%
Did the translation and cultural adaptation include changes to the text (omissions, substitutions, or additions) that could facilitate the understanding of the questions?		1		1	3	60%
Are there differences between the target and source language versions regarding the use of metaphors, idioms, and colloquialisms?		1	1	3		60%
Are the answer options the same in both language versions?				2	3	60%
If a word/sentence format or style (bold, italic, underline, etc.) was used in the source language, was the same emphasis given in the translated item?			1		4	80%
Are there any changes in score between the source and target versions that might make the item easier or harder in the translated version?		1	1	1	2	50%
Does the translated passage involve a controversial or painful topic, and is humiliating or offensive to someone?					5	100%
Do the questions include content or require skills that may be unfamiliar to respondents in either language or cultural group?	2		1	2		50%
Are there cultural differences that would have an effect on the likelihood that a response will be chosen when the item is presented in its original or translated version?	2		1	2		50%

Source: Elaborated by the authors (2022).

The Content Validity Index (CVI) was calculated for the questionnaire responses that had the highest score, as shown in the last column of Table 1. The calculations ranged from 50 to 100%.

For some authors, content validity only comprises the evaluation by a committee of experts. However, researchers have described content validity as a two-part judgmental process. The first involves the development of the instrument and, subsequently, its evaluation through analysis by specialists (Polit, 2006). To assess content validity, it is also recommended to use quantitative and qualitative procedures.

The questionnaire translated and culturally adapted into Portuguese was sent to the same two companies that translated it from English into Portuguese, but to translators who had not had contact with the first translations.

The translations and back-translations did not have significant changes. The final version translated and adapted by the specialists (Frame 1) was released for key users to access the evaluation questionnaire through Meu Clínicas©. This environment was made available to key users within a period of 7 days.

Frame 1 - The final version Portuguese Translation

Engajamento – divertido, interessante, personalizável, interativo, tem prompts (por exemplo, envia alertas, mensagens, lembretes, feedback, permite compartilhamento)

1. Interesse: O aplicativo é interessante de usar? Ele apresenta suas informações de maneira interessante em comparação com outros aplicativos semelhantes?

- 1) Nada interessante.
- 2) Desinteressante em grande parte.
- 3) OK, nem interessante nem desinteressante; envolveria o usuário por um breve período (< 5 minutos).
- 4) Moderadamente interessante; envolveria o usuário por algum tempo (5-10 minutos no total).
- 5) Muito interessante, envolveria o usuário no uso constante.

2. Entretenimento: O aplicativo é divertido/interessante de usar? Possui componente que o tornam mais divertido do que outros aplicativos semelhantes?

- 1) Chato, nada divertido ou interessante.
- 2) Entediante em grande parte.
- 3) OK, divertido o suficiente para entreter o usuário por um breve período (< 5 minutos).
- 4) Moderadamente divertido e interessante, iria entreter o usuário por algum tempo (5-10 minutos no total).
- 5) Altamente interessante e divertido, estimularia o uso repetido.
- 6) N/A Não se aplica para esse aplicativo.

3. Customização: Ele permite que você personalize as configurações e preferências que você gostaria (por exemplo, som, conteúdo e notificações)?

- 1) Não permite qualquer personalização ou requer que a configuração seja inserida todas as vezes.
- 2) Permite pouca personalização e isso limita as funções do app.
- 3) Personalização básica para funcionar adequadamente.
- 4) Permite inúmeras opções de personalização.
- 5) Permite a personalização completa das características/preferências do usuário, lembra todas as configurações.

4. Interatividade: Ele permite a entrada do usuário, fornece feedback, contém prompts (lembretes, opções de compartilhamento, notificações etc.)?

- 1) Sem recursos interativos e/ou sem resposta à entrada do usuário.
- 2) Alguns recursos interativos, mas não suficientes, que limitam as funções do aplicativo.
- 3) Recursos interativos básicos para funcionar adequadamente.
- 4) Oferece uma variedade de recursos interativos, feedback e opções de entrada do usuário.
- 5) Nível muito alto de capacidade de resposta por meio de recursos interativos, feedback e opções de entrada do usuário.

5. Grupo alvo: O conteúdo do aplicativo (visual, linguagem, design) é apropriado para o público-alvo?

- 1) Completamente inadequado, pouco claro ou confuso.
- 2) Parcialmente inadequado, pouco claro ou confuso.
- 3) Aceitável, mas não especificamente projetado para o público-alvo. Pode ser impróprio/obscuro/confuso às vezes.
- 4) Projetado para o público-alvo, com pequenos problemas.
- 5) Projetado especificamente para o público-alvo, nenhum problema encontrado.

Funcionalidade – funcionamento do aplicativo, fácil de aprender, navegação, lógica de fluxo e design gestual do aplicativo.

6. Atuação: Com que precisão/rápido os recursos do aplicativo (funções) e componentes (botões/menus) funcionam?

- 1) Aplicativo está quebrado; nenhuma/insuficiente/resposta imprecisa (por exemplo, falhas/bugs/recursos quebrados, etc.).
- 2) Algumas funções funcionam, mas estão atrasadas ou contém grandes problemas técnicos.
- 3) App funciona no geral. Alguns problemas técnicos precisam ser corrigidos ou são lentos às vezes.
- 4) Em grande parte funcional com problemas menores/insignificantes.
- 5) Totalmente funcional, nenhum bug técnico encontrado.

7. Fácil de usar: Quão fácil é aprender a usar o aplicativo; quão claras são as abas de menus, ícones e instruções?

- 1) Sem/instruções limitadas; abas de menus, ícones são confusos; complicados.
- 2) Leva muito tempo ou esforço.
- 3) Leva algum tempo ou esforço.
- 4) Fácil de aprender (ou tem instruções claras).
- 5) Capaz de usar o aplicativo imediatamente; intuitivo; simples (sem instruções necessárias).

8. Navegação: A movimentação entre as telas faz sentido; O aplicativo tem todos os links necessários entre as telas?

- 1) Nenhuma conexão lógica entre as telas/a navegação é difícil.
- 2) Compreensível após muito tempo/esforço.
- 3) Compreensível após algum tempo/esforço.
- 4) Fácil de entender/navegar.
- 5) Fluxo de tela perfeitamente lógico, fácil, claro e intuitivo por toda parte e/ou possui atalhos.

9. Design gestual: Os toques / batidas / apertos / rolagem fazem sentido? Eles são consistentes em todos os componentes/telas?

- 1) Completamente inconsistente/confuso.
- 2) Muitas vezes inconsistente/confuso.
- 3) OK, com algumas inconsistências/elementos confusos.
- 4) Principalmente consistente/intuitivo com problemas insignificantes.
- 5) Perfeitamente consistente e intuitivo.

Estética - design gráfico, apelo visual geral, esquema de cores e consistência estilística

10. Layout: A disposição e o tamanho dos botões, ícones, menus e conteúdo na tela são apropriados?

- 1) Design muito ruim, confuso, algumas opções impossíveis de selecionar, localizar, ver ou ler.
- 2) Design ruim, aleatório, pouco claro, algumas opções difíceis de selecionar/localizar/ver/ler.
- 3) Satisfatório, poucos problemas para selecionar/localizar/ver/ler itens.
- 4) Em grande parte claro, capaz de selecionar/localizar/ver/ler itens.
- 5) Profissional, simples, claro, ordenado, logicamente organizado.

11. Gráficos: Qual é a qualidade/resolução dos elementos gráficos usados para botões, ícones, menus e conteúdo?

- 1) Parecem amadores, design visual muito ruim - desproporcional estilisticamente inconsistente.
- 2) De baixa qualidade/baixa resolução; design visual de baixa qualidade – desproporcional.
- 3) De qualidade moderada e design visual (geralmente consistentes em estilo).
- 4) Elementos gráficos e design visual de alta qualidade/resolução – principalmente proporcionais consistentes em estilo.
- 5) Elementos gráficos e design visual de alta qualidade/resolução - proporcionais consistentes em estilo por toda parte.
- 6) N/A Não há informações dentro do aplicativo.

12. Apelo visual: Quão bom é o aplicativo?

- 1) Feio, desagradável de se ver, mal projetado, conflitante, cores incompatíveis.
- 2) Ruim – mal projetado, mau uso de cores, visualmente chato.
- 3) OK – mediano, nem agradável, nem desagradável.
- 4) Agradável – gráficos perfeitos – consistente e projetados profissionalmente.
- 5) Bonito – muito atraente, memorável, se destaca; o uso de cores aprimora os recursos/menus do aplicativo.

Informações – Contém informações de alta qualidade (por exemplo, texto, feedback, medidas, referências) de uma fonte confiável

13. Qualidade da informação: O conteúdo do aplicativo está correto, bem escrito e relevante para o objetivo/tópico do aplicativo?

- 1) Irrelevante/inadequado/incoerente/incorreto.
- 2) Pobre. Pouco relevante/apropriado/coerente/pode estar incorreto.
- 3) Moderadamente relevante/apropriado/coerente/e parece correto.
- 4) Relevante/adequado/coerente/correto.
- 5) Altamente relevante, apropriado, coerente e correto.
- 6) N/A Não há informações no aplicativo.

14. Quantidade de informações: As informações no aplicativo são completas e objetivas?

- 1) Mínima ou com muita informação.
- 2) Insuficiente ou possivelmente completas e objetivas.
- 3) As informações são OK.
- 4) Oferece uma ampla gama de informações, possui algumas lacunas ou detalhes desnecessários; ou não tem links para mais informações e recursos.
- 5) As informações são completas e objetivas; contém links para mais informações e recursos.
- 6) N/A Não há informações no aplicativo.

15. Informações visuais: A explicação visual dos conceitos – através de tabelas/gráficos/imagens/vídeos, etc. – é clara, lógica, correta?

- 1) Nada claro/confuso/errado ou necessário, mas ausente.
- 2) Geralmente pouco claro/confuso/errado.
- 3) Aceitável, porém às vezes pouco claro/confuso/errado.
- 4) Em grande parte claro/lógico/correto.
- 5) Perfeitamente claro/lógico/correto.
- 6) N/A Não há informações visuais no aplicativo (por exemplo, contém apenas áudio ou texto).

16. Credibilidade da fonte: as informações dentro do aplicativo parecem vir de uma fonte confiável?

- 1) Fonte suspeita.
- 2) Falta credibilidade.
- 3) Não é suspeito, mas a legitimidade da fonte não é clara.
- 4) Possivelmente vem de uma fonte legítima.
- 5) Definitivamente vem de uma fonte legítima/especializada.
- 6) N/A Não há informações no aplicativo.

17. Você recomendaria este aplicativo para pessoas que possam se beneficiar dele?

- 1) Não recomendo este aplicativo a ninguém.

- 2) Há muito poucas pessoas para quem eu recomendaria este aplicativo.
 - 3) Talvez existam várias pessoas para quem eu recomendaria este aplicativo.
 - 4) Há muitas pessoas para quem eu recomendaria este aplicativo.
 - 5) Definitivamente eu recomendaria este aplicativo a todos.
18. Quantas vezes você acha que usaria este aplicativo nos próximos 12 meses se fosse relevante para você?
- 1) Nenhuma, pois talvez não seja necessário; 2) 1-2 vezes, se necessário; 3) 3-10 vezes, se necessário; 4) 10-50 vezes; 5) >50 vezes.
19. Você pagaria por este aplicativo?
- 1) Definitivamente não; 2) Acredito que não; 3) Não sei; 4) Acredito que sim; 5) Definitivamente sim.
20. Qual é a sua classificação geral (estrelas) do aplicativo?
- 1) * Um dos piores aplicativos que já usei
 - 2) **
 - 3) *** Média
 - 4) ****
 - 5) ***** Um dos melhores aplicativos que usei Impacto percebido
21. Conscientização – Este aplicativo aumentou minha conscientização sobre a importância de abordar meus hábitos/comportamentos saudáveis.
- 1) Discordo totalmente; 2) Discordo em algumas coisas; 3) Não concordo e nem discordo; 4) Concordo em algumas coisas; 5) Concordo totalmente.
22. Conhecimento– Este aplicativo aumentou meu conhecimento/compreensão para meus hábitos/comportamentos saudáveis.
- 1) Discordo totalmente; 2) Discordo em algumas coisas; 3) Não concordo e nem discordo; 4) Concordo em algumas coisas; 5) Concordo totalmente.
23. Atitudes – O aplicativo mudou minhas atitudes para melhorar esse meus hábitos/comportamentos saudáveis.
- 1) Discordo totalmente; 2) Discordo em algumas coisas; 3) Não concordo e nem discordo; 4) Concordo em algumas coisas; 5) Concordo totalmente.
24. Intenção de mudar – O aplicativo aumentou minhas intenções/motivação para abordar hábitos/comportamentos saudáveis.
- 1) Discordo totalmente; 2) Discordo em algumas coisas; 3) Não concordo e nem discordo; 4) Concordo em algumas coisas; 5) Concordo totalmente.
25. Busca de ajuda – Este aplicativo me encorajaria a procurar mais ajuda para lidar com meus hábitos/comportamentos saudáveis (se eu precisar).
- 1) Discordo totalmente; 2) Discordo em algumas coisas; 3) Não concordo e nem discordo; 4) Concordo em algumas coisas; 5) Concordo totalmente.
26. Mudança de comportamento – O uso deste aplicativo aumentará/diminuirá meus hábitos/comportamentos saudáveis.
- 1) Discordo totalmente; 2) Discordo em algumas coisas; 3) Não concordo e nem discordo; 4) Concordo em algumas coisas; 5) Concordo totalmente.
27. Mais comentários sobre o aplicativo?
Questão aberta discursiva.

Source: Elaborated by the authors (2022).

The invitation to participate in the survey was made available on October 3, 2022, and users had a period of 5 weeks to participate. The survey was disabled for viewing in the application on November 7th of the same year. The sample for the validation stage of this study consisted of 107 participants. The age of participants ranged from 18 years to 65 years or older. Most participants were concentrated in the 35 to 44 age group (31.5%), followed by 24.1% in the 45 to 54 age group and 19.4% of participants aged 55 to 64 years answered the search. Only one application user aged up to 18 answered the questionnaire. Regarding the level of education, in this study, 37% of the participants finished higher education and 31.5% had completed high school. Almost 10% of the respondents have complete or incomplete primary and secondary education and 16.7% answered that they have incomplete higher education.

The results for calculating the reliability of the questionnaire are presented according to the uMARS divisions. The first part of the instrument presented a Cronbach's alpha index of 0.70, as shown in Table 2. The minimum acceptable value for the

reliability of a questionnaire, among most authors, is $\alpha \geq 0.60$ (below this value the internal consistency of the scale used is considered low) and the maximum expected value is $\alpha = 0.90$.

Table 2 - Confiabilidade respostas parte 1 – Engajamento.

Engagement	Total sum	Average	Sum of Variance	Variance
Interest	458	4,28	77,59	0,73
Entertainment	296	2,76	351,16	3,34
Customization	261	2,51	157,99	1,53
Interactivity	314	2,96	115,85	1,10
Target group	447	4,26	78,06	1,75
Cronbach's alpha index: 0,70				

Source: Elaborated by the authors (2022).

Table 3 shows the questions that increased the questionnaire's reliability, with a reliability level of 0.83.

Table 3 - Reliability answers part 2 – Functionality.

Functionality	Total sum	Average	Sum of Variance	Variance
Performance	422	3,98	111,96	1,07
Easy to use	437	4,08	72,24	0,68
Navigation	411	3,95	54,76	0,53
Gestural Design	442	4,25	85,56	0,83
Cronbach's alpha index: 0,83				

Source: Elaborated by the authors (2022).

The reliability of the questionnaire can be reduced the smaller the number of items. This result is shown in Table 4, where there was a reduction in the reliability index to 0.60 due to the low number of items to be measured.

Table 4 - Aesthetics answers part 3 – Aesthetics

Aesthetics	Total sum	Average	Sum of Variance	Variance
Layout	427	4,03	104,92	1,00
Graphics	399	3,8	222,36	2,14
Visual appeal	411	3,95	56,76	0,55
Cronbach's alpha index: 0,60				

Source: Elaborated by the authors (2022).

It is important to point out that, although the scientific literature on the applications of the α coefficient in different areas of knowledge is wide and comprehensive, there is still no consensus among researchers about the interpretation of the reliability of a questionnaire obtained from the value of this coefficient. In general, a survey instrument that obtains $\alpha \geq 0.70$ is considered satisfactory. Table 5 presents the calculation of the 0.83 coefficient for the sub-scale related to evaluations of the application's information.

Table 5 - Information answers part 4 – Information.

Information	Total sum	Average	Sum of Variance	Variance
Information quality	400	3,70	202,80	1,90
Amount of information	428	4,07	135,39	1,30
Visual information	413	3,97	190,06	1,85
Source credibility	473	4,50	102,25	0,98
Cronbach's alpha index: 0,83				

Source: Elaborated by the authors (2022).

It is important to highlight that people generally have different perceptions and interpretations of the items being evaluated. Thus, considering a significant sample of participants in the survey, it is very likely that there is no certain agreement regarding the response to a given item. In this way, the Cronbach coefficient of 0.69 for Table 6 is justified.

Table 6 - Reliability answers part 5 – Subjective Quality.

Subjective Quality	Total sum	Average	Sum of Variance	Variance
Recommendation	454	4,24	81,68	0,77
Utility	381	3,56	97,68	0,92
Investment	256	2,39	145,79	1,38
Classification	411	3,84	73,54	0,69
Cronbach's alpha index: 0,68				

Source: Elaborated by the authors (2022).

The evaluation by parts showed that five of its domains presented an alpha between 0.6 and 0.83. According to Jeannie Sneed and Herman (1990), this value indicates that there is a satisfactory correlation between the items and demonstrates that the instrument has good reliability. The last part, on specific health assessment, was the one with the best internal consistency (1.01), and parts 3 (Aesthetics) and 5 (Subjective quality) were the ones with the lowest value (0.63 – 0.68).

From the calculation of the average of the coefficients, the results of this study show that the instrument of this research has a Cronbach coefficient of 0.78. Thus, it is possible to state that the questionnaire has the desired consistency in all blocks. It proved to be efficient for the research, being, therefore, usable in future applications.

The translation and validation process is important, a literal translation could lead to the development of an instrument that is not equivalent to the original questionnaire. As well as, the formation of a commission of specialists is of paramount importance to guarantee a cross-cultural equivalence.

The back translations were similar to each other and equivalent to the original version of uMARS in almost all of its items. Some adaptations were made by experts, but they maintained an agreement with the original version. In addition, some responses were maintained by the expert committee, as they did not believe that the meaning of the original instrument was changed.

After validation and consent of the translations, the experts answered a questionnaire with 10 questions for content validation and semantic evaluation of the questionnaire. The Content Validity Index (CVI) was calculated for each response to the questionnaire applied to the experts. Responses ranged from completely agree and partially agree between 50% to 100% agreement.

According to Domnich et al. (2016), translating IT terminologies, full of anglicisms, can be a challenge in areas not related to IT, such as public health and health. Unlike some other languages, Italian tends to absorb IT terminology in an unadapted way. Before starting the first stage of translation, local websites that use mobile technologies were consulted, including press releases from major smartphone manufacturers, and several discrepancies were observed in the translation of terms with the same semantic-referential meanings. The critical evaluation of these semantic neologisms helped to achieve linguistic equivalence between the source and target languages.

Both for the MARS and for its translated versions and for the questionnaire provided to the participants, it is recommended to watch a video on technical explanations of the content of the questions, before answering the questionnaire. However, since uMARS is a more simplified version, aimed at users of the application, an explanatory video is not necessary.

According to the authors who evaluated the effectiveness of the evaluation instrument through the MARS and uMARS questionnaires, the notes referring to the general classification given to the application differ from those presented in the application stores. The same occurs in this study, currently, Meu Clinicas© has a rating of 3.3 stars, an average of stars based on

the vote of 209 users, with the largest volumes of ratings being 5 stars and 1 star. According to the app store, more than 10,000 downloads of the tool have already been carried out. However, the assessment instrument translated into Portuguese presented an overall rating of 3.8 stars for this research. The results showed that only one participant classified the application with only 1 star, that is, as a bad application to use. However, most participants (45.3%) gave the application four stars, followed by three stars with 27.4% and 23.6% gave the application the maximum score of 5 stars.

To calculate Cronbach's alpha coefficient, the correlation between the instrument's responses was evaluated by the average correlation between the questions. This calculation is performed by analyzing the profile of responses provided by research participants using an equation that uses the variance of individual items and the variance of the sum of items for each participant. Internal consistency analysis provided a Cronbach's alpha value of 0.78 — an acceptable value. The main method used to establish the internal consistency of the instruments is the calculation of Cronbach's alpha coefficient, which estimates how uniformly the items contributed to the unweighted sum of the instrument (Pasquali, 2017; Pasquali, 2007).

The uMARS total score showed excellent internal consistency (Cronbach's alpha = 0.90). The internal consistencies of their subscales were also very high (alpha of engagement = 0.80; alpha of functionality = 0.70; alpha of aesthetics = 0.71; alpha of information = 0.78; and alpha of satisfaction = 0.78) (Stoyanov et al., 2016). According to Terwee et al. (2007), alpha values should be between 0.7 and 0.9. Values lower than 0.7 indicate that there is a low correlation between the items and values greater than 0.9 indicate a very strong correlation, which suggests the existence of redundancy between the items. The alpha must be calculated for the scale as a whole and for each domain.

4. Conclusion

The Portuguese version of uMARS was produced, a valid and reliable tool to evaluate health-related apps available in app stores. In the uMARS validation process, all steps were carried out in a methodologically careful and controlled manner, and do not differ from the results found in other validation studies of this scale carried out in other countries.

As a limiting factor of this research, it was not possible to carry out other types of validation, such as, for example, the clinical validation of the patients who answered the questionnaire.

It is recommended that, in a future study, the uMARS in the Portuguese version be applied to other health-related applications in different regions of the country.

Therefore, this work presents an important contribution to the health area by making available the Brazilian version of a valid and reliable questionnaire to evaluate mobile applications in the health area aimed at patients.

References

- Alexandre, N. M. C., & Coluci, M. Z. O. (2011). Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. *Ciencia & saude coletiva*, 16, 3061-3068.
- Bardus, M., Awada, N., Ghandour, L. A., Fares, E. J., Gherbal, T., Al-Zanati, T., & Stoyanov, S. R. (2020). The Arabic version of the Mobile App Rating Scale: development and validation study. *JMIR mHealth and uHealth*, 8(3), e16956.
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186-3191.
- Bermudes, W. L., Santana, B. T., Braga, J. H. O., & Souza, P. H. (2016). Tipos de escalas utilizadas em pesquisas e suas aplicações. *Revista Vértices*, 18(2), 7-20.
- Clason, D. L., & Dormody, T. J. (1994). Analyzing data measured by individual Likert-type items. *Journal of agricultural education*, 35(4), 4.
- Davalbhakta, S., Advani, S., Kumar, S., Agarwal, V., Bhojar, S., Fedirko, E., ... & Agarwal, V. (2020). A systematic review of smartphone applications available for corona virus disease 2019 (COVID19) and the assessment of their quality using the mobile application rating scale (MARS). *Journal of medical system*, 44, 1-15.

- Domnich, A., Arata, L., Amicizia, D., Signori, A., Patrick, B., Stoyanov, S., ... & Panatto, D. (2016). Development and validation of the Italian version of the Mobile Application Rating Scale and its generalisability to apps targeting primary prevention. *BMC Medical Informatics and Decision Making*, 16(1), 1-10.
- GIL, Antonio. C. *Como elaborar projetos de pesquisa*. São Paulo: Atlas, ed. 4, 2002.
- LeBeau, K., Huey, L. G., & Hart, M. (2019). Assessing the quality of mobile apps used by occupational therapists: evaluation using the user version of the mobile application rating scale. *JMIR mHealth and uHealth*, 7(5), e13019.
- Menegotto, A. B., Pereira, G. M., Rocha, J. F., & Malvezzi, M. L. F. (2019). Meu Clínicas®, aplicativo para pacientes: inovação em tecnologia da informação e comunicação (TIC) na saúde pública. *Clinical and biomedical research*. Porto Alegre.
- Melin, J., Bonn, S. E., Pendrill, L., & Lagerros, Y. T. (2020). A questionnaire for assessing user satisfaction with mobile health apps: development using Rasch measurement theory. *JMIR mHealth and uHealth*, 8(5), e15909.
- Messner, E. M., Terhorst, Y., Barke, A., Baumeister, H., Stoyanov, S., Hides, L., ... & Probst, T. (2020). The German version of the Mobile App Rating Scale (MARS-G): development and validation study. *JMIR mHealth and uHealth*, 8(3), e14479.
- Payo, R. M., Álvarez, M. F., Díaz, M. B., Izquierdo, M. C., Stoyanov, S. R., & Suárez, E. L. (2019). Spanish adaptation and validation of the Mobile Application Rating Scale questionnaire. *International Journal of Medical Informatics*, 129, 95-99.
- Pasquali, L. (2017). *Psicometria: teoria dos testes na psicologia e na educação*. Editora Vozes Limitada.
- Pasquali, L. (2007). Validade dos testes psicológicos: será possível reencontrar o caminho? *Psicologia: teoria e pesquisa*, 23, 99-107.
- 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.
- Richardson, R. J. et al. (1999) *Pesquisa social: métodos e técnicas*. Atlas, (3a ed.).
- Souza, A. C. D., Alexandre, N. M. C., & Guirardello, E. D. B. (2017). Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. *Epidemiologia e serviços de saúde*, 26, 649-659.
- Jeannie Sneed PhD, R. D., & Herman, C. M. (1990). Influence of job characteristics and organizational commitment on job satisfaction of hospital foodservice employees. *Journal of the American Dietetic Association*, 90(8), 1072-1076.
- 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), e3422.
- Stoyanov, S. R., Hides, L., Kavanagh, D. J., & Wilson, H. (2016). Development and validation of the user version of the Mobile Application Rating Scale (uMARS). *JMIR mHealth and uHealth*, 4(2), e5849.
- Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., ... & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*, 60(1), 34-42.
- Vagias, W. M. (2006). Likert-type scale response anchors. Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University.
- Waltz, C. F., & Bausell, B. R. (1981). *Nursing research: design statistics and computer analysis*. Davis Fa.
- Zhou, L., Bao, J., Setiawan, I. M. A., Saptono, A., & Parmanto, B. (2019). The mHealth App Usability Questionnaire (MAUQ): development and validation study. *JMIR mHealth and uHealth*, 7(4), e11500.