# Translation and validation of the Italian version of the user version of the Mobile Application Rating Scale (uMARS)

SIMONE MORSELLI[1,2], ARCANGELO SEBASTIANELLI[1,2], ALEXANDER DOMNICH[3], CHIARA BUCCHI[1],
PIETRO SPATAFORA[1,2], ANDREA LIACI[1,2], LUCA GEMMA[1,2], STAVROS GRAVAS[4], DONATELLA PANATTO[5],
STOYAN STOYANOV[6], SERGIO SERNI[1,2], MAURO GACCI[1,2]

[1] Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy; [2] Unit of Minimally Invasive and Robotic Urologic Surgery and Kidney Transplantation, Careggi University Hospital, Florence, Italy; [3] Hygiene Unit, San Martino Policlinico Hospital, IRCCS for Oncology and Neurosciences, Genoa, Italy; [4] Department of Urology, University of Thessaly, Larissa, Greece; [5] Department of Health Sciences, University of Genoa, Genoa, Italy; [6] School of Design, Creative Industries Faculty, Queensland University of Technology, Brisbane, QLD, Australia

## Summary

**Background**. *Health sciences are steadily developing apps to help people to adopt correct lifestyles and to help physicians to monitor patients with chronic diseases. However, a properly validated tool that can evaluate patients' perception of apps is still lacking in many languages. In English, a validated questionnaire, called User Version of the Mobile Application Rating Scale (uMARS), is currently available. We translated the uMARS into Italian and validated our version.*

**Methods**. *The uMARS questionnaire was translated from English to Italian by an official translator, and then administered to 100 smartphone users in order to evaluate the same app at times 1 and 2 (after 2 weeks). Paired t-test, Pearson Correlation Coefficient, Intraclass Correlation Coefficient (ICCs) and Cronbach's Alpha were used to evaluate the reliability and validity of the Italian uMARS.*

**Results**. *We recruited 100 subjects, 52 males (52%) and 48 females (48%), with a mean age of 22.8 (SD: 3.4). All subjects answered all questions both at time 1 and at time 2. Paired t-test showed no statistically significant difference in each answer or group of answers between times 1 and 2 (P > 0.05). Cronbach's alpha was 0.945, as all subjects answered all questions. Each question was further assessed through the Pearson correlation coefficient, which demonstrated high reliability, with significant P (< 0.05) and Pearson Coefficients higher than 0.7. Similarly, ICCs were always higher than 0.750.*

**Conclusions**. *Our results validated the Italian version of uMARS, which may become a reliable and useful tool for evaluating health apps.*

## Introduction

Mobile health (mHealth) is a continuously evolving field of healthcare, owing to its particularly attractive features, such as ubiquity and portability [1]. Indeed, the number of mobile subscriptions in Europe is over 1,300 per 1,000 inhabitants, and 76% of Europeans access the Internet daily [2]. Healthcare professionals (HCPs), patients and the general public are increasingly using mHealth for a vast range of purposes, including communication and consultation with health services, HCPs and patients, the acquisition of health information and monitoring [1, 3].

Today, health-related mobile applications (apps) constitute a cornerstone of mHealth, and their overall market in 2019 was estimated as > 350,000 [4]. Indeed, it is believed that more than 500 million people worldwide have downloaded at least one mHealth app to their mobile phone [5].

Despite this promising outlook, mHealth apps are often of suboptimal quality and,worryingly, have poor or no evidence base [6]. This fact may compromise users'

choices and habits, thereby affecting health-related outcomes [7]. The easiest way to assess the quality of an app is to check its "star rating", which can immediately be done in app stores. However, this quality appraisal system is highly subjective and may be significantly skewed by the phenomenon of "information asymmetry" between specialists and lay users [8, 9]. For instance, a review of urology apps available in the common app stores demonstrated that, while some apps were of moderate-high quality, the average app-store rating was 1 star only [10].

In order to address the above-described quality issues in a more objective way, the Mobile Application Rating Scale (MARS) was developed [8]. Given the frequency of its citation in published papers), MARS is probably the most popular scale worldwide. The original English language MARS scale has already been successfully validated in several languages, including Italian, Spanish, German, Dutch and Arabic [9, 11-14]. Briefly, the original MARS instrument is composed of 23 Likert scale-based items that cover the objective quality

dimensions of engagement, functionality, esthetics and information, and one subjective dimension [8].

However, the original MARS scale was developed for professional use by researchers, clinicians and other specialists, and training is required before it can be used [8]. Given that most app users have no specialist knowledge, a training-free user version of MARS (uMARS) was subsequently validated and is publicly available [15].

An Italian version of the "expert" MARS instrument was successfully validated by Domnich et al. (2016). The same research group then tried to adapt the Italian version of MARS to the user version, with regard to the assessment of the quality of an app concerning invasive pneumococcal disease; however, no formal validation was performed [16]. To address this unmet need, the present study aimed to validate the Italian version of uMARS.

# Methods

### The English version of uMARS

Like the expert version, the uMARS covers four objective dimensions (engagement, functionality, esthetics and information) and one subjective dimension. Unlike the expert version, however, it is training-free, omits three items on the information subscale and has better readability properties. Briefly, the scale consists of 20 anchored 5-point Likert-type items that are distributed as follows: engagement (N = 5), functionality (N = 4), esthetics (N = 3) and information (N = 4) and 4 items belonging to the subjective quality domain. There is also a section on perceived impact (6 items), which assesses the potential impact of a given app on users' knowledge, behaviors, intentions, help-seeking, etc.; this is a domain that evaluates users' perception of the usefulness of the app, thereby enabling its impact to be rated. The scoring procedure is based on the mean: (i) subscale-specific scores are obtained by averaging individual item scores on that particular subscale; (ii) an overall uMARS score is an average of the four objective dimensions [15].

During the validation process, the English version of uMARS displayed both excellent internal consistency (Cronbach's α of 0.90) and good test-retest reliability in a period of up to three months [15].

### Adaptation and translation

To meet our objective, we generally adopted the process of adaptation/translation of the Italian version of MARS [9]. Following consultations with the authors of the expert MARS version, it was deemed that the two Italian versions were fully interchangeable from the points of view of conceptual, item, semantic, operational, measurement and functional equivalence. Indeed, the uMARS has simpler wording, but at the same time shares the same topics with the expert version. The formal adaptation procedure was therefore judged redundant.

The English version of uMARS was then translated into Italian by a professional bilingual translator. The resulting output was then compared with the validated expert scale and discussed by the research group, which led to only minor changes being made. A back-translation was then performed in order to verify the compatibility and accuracy of meaning between the source and target languages. The authors of the original Italian MARS scale approved the final Italian uMARS questionnaire [9]. The final version of the questionnaire is available in Appendix A.

### Validation procedure

To validate the final Italian uMARS scale, we roughly followed the original methodology, as described by Stoyanov et al. (2016). We aimed to test the internal consistency and test-retest reliability of the scale.

To validate the questionnaire, we set a target of 100 individuals, a sample size in line with the original uMARS validation study [15]. Potential participants were selected from the University of Florence, Faculty of Medicine. Subjects who, on the day of enrollment, were able to navigate in the study target app from their devices were potentially eligible. Participants had to be sufficiently fluent in Italian.

The app chosen for uMARS validation was Facebook. There were several strong arguments for this choice: this app is free, popular and does not require participants to spend the time they would normally need to familiarize themselves with a new app, and Facebook contains all of the components covered by the four uMARS domains; therefore, all items could be confidently validated. The choice of Facebook was also related to the cost of downloading some health apps and the unavailability of some health apps on various mobile operating systems. Moreover, the use of certain health apps would have excluded some subjects from participating in the study.

Eligible subjects were instructed to navigate all the app functions for at least 10 min and then to rate the app by using the uMARS. In order to ascertain intra-rater reproducibility, this procedure was carried out twice: on enrollment (t1) and approximately two weeks later (t2); this time-lag was deemed an appropriate waiting period by the research team.

The study was conducted in accordance with the Declaration of Helsinki. Participation in this research was voluntary and all relevant Italian and international biomedical and privacy-related guidelines were followed. As the nature of the study was neither interventional nor biomedical, participants were not exposed to any risks. Therefore, formal ethical approval for this study was deemed unnecessary. Furthermore, all participants were enrolled on a voluntary basis, and our research group guaranteed their privacy.

### Statistical analysis

Internal consistency was measured by means of Cronbach's α using the following "rule of thumb" categories: excellent (≥ 0.90), good (0.80-0.89),

acceptable (0.70-0.79), questionable (0.60-0.69), poor (0.50-0.59) and unacceptable (< 0.50) [17].

The reproducibility of the Italian uMARS questionnaire was evaluated through the second application of the questionnaire at t2, two weeks after t1. First, a paired t-test was calculated in order to observe variations between the two measurements, for each sub-score and total score. Moreover, for every variable, test-retest reliability was evaluated through a bivariate Pearson correlation coefficient. A Pearson correlation coefficient above 0.7 and a significance level below 0.05 were considered sufficient to assess test-retest reliability. Finally, intraclass correlation coefficients (ICCs) were computed between t1 and t2, in order to provide weighted values of rater agreement and assess proximity rather than equality of ratings. The model chosen to calculate ICCs was that of random-effects average measures with absolute agreement [18].

All statistical analyses were conducted by means of SPSS version 20.0 (SPSS Inc, Chicago, IL, USA). A P-value < 0.05 was conventionally deemed statistically significant.

Tab. I. Characteristics of participants in the validation of the Italian version of uMARS.

| Sex, n (%) | Female | 48 (48.0%) |
|---|---|---|
| | Male | 52 (52.0%) |
| Native Italian speaker, n (%) | | 100 (100%) |
| Education, n (%) | High school | 88 (88.0%) |
| | University | 9 (9.0%) |
| | Middle school | 3 (3.0%) |
| Age, mean (SD) | | 22.8 (3.4) |

Tab. II. Test-retest reliability between time 1 and time 2 in the validation of the Italian version of uMARS.

| | | Paired t-test | | | Reliability | |
|---|---|---|---|---|---|---|
| | | Time 1 | Time 2 | P | Pearson correlation | P |
| A - engagement | A1 - entertainment | 3.8 (1.0) | 3.7 (1.1) | 0.235 | 0.826 | 0.001 |
| | A2 - interest | 3.9 (1.1) | 3.8 (1.1) | 0.352 | 0.810 | 0.001 |
| | A3 - customization | 4.0 (0.9) | 3.8 (0.9) | 0.309 | 0.810 | 0.001 |
| | A4 - interactivity | 3.9 (0.9) | 4.0 (0.9) | 0.498 | 0.797 | 0.001 |
| | A5 - target group | 3.9 (1.0) | 3.8 (1.1) | 0.822 | 0.798 | 0.001 |
| A - engagement total | 19.4 (3.6) | 19.1 (3.8) | 0.283 | 0.854 | 0.001 | |
| B - functionality | B6 - performance | 4.0 (0.9) | 4.0 (0.8) | 0.750 | 0.828 | 0.001 |
| | B7 - ease of use | 4.0 (0.9) | 3.9 (0.9) | 0.623 | 0.742 | 0.001 |
| | B8 - navigation | 4.1 (0.7) | 3.9 (0.8) | 0.160 | 0.803 | 0.001 |
| | B9 - gestural design | 4.2 (0.8) | 4.2 (0.6) | 0.660 | 0.843 | 0.001 |
| B - functionality total | 16.3 (2.1) | 16.0 (2.0) | 0.214 | 0.790 | 0.001 | |
| C - esthetics | C10 - layout | 3.8 (0.8) | 3.9 (0.8) | 0.743 | 0.748 | 0.001 |
| | C11 - graphics | 3.9 (0.8) | 3.8 (0.9) | 0.822 | 0.771 | 0.001 |
| | C12 - visual appeal | 3.8 (0.9) | 3.8 (1.0) | 0.323 | 0.802 | 0.001 |
| C - esthetics total | 11.5 (2.0) | 11.4 (2.1) | 0.727 | 0.794 | 0.001 | |
| D - information | D13 - quality of information | 3.7 (1.0) | 3.7 (0.8) | 0.643 | 0.753 | 0.001 |
| | D14 - quantity of information | 3.9 (1.1) | 3.9 (1.0) | 0.599 | 0.850 | 0.001 |
| | D15 - visual information | 4.0 (0.9) | 3.9 (1.0) | 0.323 | 0.772 | 0.001 |
| | D16 - credibility of source | 3.6 (1.1) | 3.6 (1.1) | 1.000 | 0.844 | 0.001 |
| D - information total | 15.1 (3.4) | 15.1 (3.1) | 1.000 | 0.874 | 0.001 | |
| Quality | (A+b+c+d/4) | 15.6 (2.3) | 15.4 (2.3) | 0.339 | 0.881 | 0.001 |
| E - subjective quality | E17 - recommendation to others | 3.7 (1.0) | 3.7 (1.0) | 0.700 | 0.868 | 0.001 |
| | E18 - use and relevance | 4.1 (1.1) | 3.9 (1.1) | 0.167 | 0.792 | 0.001 |
| | E19 - payment | 2.5 (1.6) | 2.5 (1.5) | 0.686 | 0.803 | 0.001 |
| | E20 - overall rating | 3.7 (0.8) | 3.5 (1.0) | 0.127 | 0.792 | 0.001 |
| E - subjective quality total | 14.0 (3.2) | 13.6 (3.2) | 0.080 | 0.880 | 0.001 | |
| F - perceived impact | F1 - awareness | 3.3 (1.2) | 3.2 (1.1) | 0.294 | 0.810 | 0.001 |
| | F2 - knowledge | 3.0 (1.2) | 3.1 (1.3) | 0.230 | 0.868 | 0.001 |
| | F3 - attitudes | 3.0 (1.3) | 3.0 (1.2) | 0.822 | 0.849 | 0.001 |
| | F4 - intention to change | 3.3 (1.1) | 3.1 (1.2) | 0.128 | 0.813 | 0.001 |
| | F5 - help seeking | 3.2 (1.2) | 3.2 (1.2) | 0.538 | 0.800 | 0.001 |
| | F6 - behavior change | 3.1 (1.3) | 3.2 (1.2) | 0.800 | 0.886 | 0.001 |
| F - perceived impact total | 19.0 (6.4) | 18.8 (6.3) | 0.488 | 0.950 | 0.001 | |
| Total score | 83.8 (15.4) | 82.5 (14.8) | 0.197 | 0.917 | 0.001 | |

icc: intraclass correlation coefficient; for single measures α = cronbach's alpha. All values are scores and are reported as mean (standard deviation). Paired t-test was done as statistical analysis between time 1 and time 2.

## Results

During the study, the target population of 100 was reached in approximately three months. The mean age of participants was 22.8 years (SD = 3.4) and the male-to-female ratio was close to 1 (48 female *vs* 52 male). All participants were native Italian speakers, mostly university students in their 2nd, 3rd or 4th year; 22 participants (22%) were students' relatives: 9 (9%) with a degree, 10 (10%) with a high-school diploma and 3 (3%) without a high-school diploma. All participants filled in the uMARS questionnaire both at t1 and at t2 (100% compliance). Their principal characteristics are reported in Table I.

The Italian uMARS displayed high internal consistency (Cronbach's α = 0.95), and excellent reliability (Cronbach's α based on standardized items = 0.98).

The paired t-tests demonstrated that all 20 items and all the overall (sub)scale scores were similar between t1 and t2 (P > 0.12) (Table II). In detail, the highest variability was observed for question E20, regarding the subjective evaluation of the app: "What is your overall (star) rating of the app?" t1 mean 3.7 (SD: 0.8) *vs* t2 mean 3.6 (SD: 1.0) p = 0.089. The lowest variability was observed in the Information Section concerning question D16 "Credibility of source: does the information within the app seem to come from a credible source?", t1 mean 3.6 (SD: 1.1) t2 mean 3.6 (SD: 1.1) p = 1.000. Moreover, test-retest reliability showed high consistency in each sub-score and total score, with a Pearson Correlation coefficient above 0.7 and significance below 0.05 (Tab. II). Analogously, the ICCs observed were constantly high (Tab. III), confirming excellent test-retest reliability.

Tab. III. Intraclass Correlation Coefficients between Time 1 and Time 2 in the validation of the Italian version of uMARS.

| | | ICC | α |
|---|---|---|---|
| A- engagement | A1 - entertainment | 0.85 | 0.97 |
| | A2 - interest | 0.83 | 0.90 |
| | A3 - customization | 0.81 | 0.90 |
| | A4 - interactivity | 0.80 | 0.97 |
| | A5 - target group | 0.78 | 0.91 |
| A - engagement total | 0.80 | 0.89 | |
| B - functionality | B6 - performance | 0.82 | 0.97 |
| | B7 - ease of use | 0.76 | 0.87 |
| | B8 - navigation | 0.80 | 0.87 |
| | B9 - gestural design | 0.87 | 0.94 |
| B - functionality total | 0.79 | 0.88 | |
| C - esthetics | C10 - layout | 0.75 | 0.87 |
| | C11 - graphics | 0.77 | 0.90 |
| | C12 - visual appeal | 0.80 | 0.90 |
| C - esthetics total | 0.79 | 0.88 | |
| D - information | D13 - quality of information | 0.76 | 0.85 |
| | D14 - quantity of information | 0.85 | 0.92 |
| | D15 - visual information | 0.77 | 0.87 |
| | D16 - credibility of source | 0.84 | 0.95 |
| D - information total | 0.87 | 0.93 | |
| Quality | (A+b+c+d/4) | 0.88 | 0.94 |
| E -subjective quality | E17 - recommendation to others | 0.87 | 0.92 |
| | E18 - use and relevance | 0.79 | 0.86 |
| | E19 - payment | 0.80 | 0.85 |
| | E20 - overall rating | 0.81 | 0.87 |
| E - subjective quality total | 0.88 | 0.94 | |
| F – perceived impact | F1 - awareness | 0.81 | 0.87 |
| | F2 - knowledge | 0.86 | 0.93 |
| | F3 - attitudes | 0.85 | 0.92 |
| | F4 - intention to change | 0.81 | 0.90 |
| | F5 - help seeking | 0.80 | 0.89 |
| | F6 - behavior change | 0.88 | 0.94 |
| F- perceived impact total | 0.95 | 0.97 | |
| Total score | 0.92 | 0.96 | |

ICC: Intraclass Correlation Coefficient; for single measures α = Cronbach's Alpha. All values are scores and are reported as mean (standard deviation). Paired t-test was done as statistical analysis between time 1 and time 2.

## Discussion

The uMARS scale is designed to enable the quality of health-related apps to be evaluated, both within institutional bodies (e.g. researchers, scientific societies, regulatory agencies, industry) and among end-users [15]. We translated this scale into Italian and validated our version. Together with the Italian "expert" MARS questionnaire, our newly validated Italian uMARS scale completes the set of instruments for the quality evaluation of Italian health apps. Although Italy ranks relatively highly in terms of mobile subscriptions and Internet access, the overall English language proficiency of Italians is among the lowest in Europe [9, 19-21]. Therefore, original questionnaires should be translated into Italian and validated, in order to reach the majority of the population (this is also a matter of equality of access to a new technology). Moreover, the Italian version of uMARS can now also be used for wide population-based research.

The validity and reliability of the Italian uMARS version proved similar to those of the original uMARS [15]. Specifically, we obtained a similarly high level of internal consistency ($\alpha \geq 0.90$), in line with the Italian "expert" version of MARS [10]. The test-retest reliability and ICCs were also high, being similar to those of the original scale created by Stoyanov et al. (2016). In our opinion, these optimal properties were obtained because: (i) we adopted the methodology of the original MARS and consulted the original uMARS and the Italian MARS scales, and (ii) subsequent approval of the text of the scale was obtained from the authors of the previously published "expert" scale [8, 9, 15].

Considering that the present study was based on a body of previous research, we were able to identify some limitations of our study. Indeed, the total sample size of 100 participants was not powered *a priori* [8, 9, 15]. However, in a recently published review by Bujang et al., this sample size was judged sufficient [22]. Moreover, our sample was mostly composed of young university students, who are usually very familiar with modern technology and use apps more often than older people do. Consequently, the sample might be not fully representative of the general Italian population.

Finally, although the app chosen for validation had the advantage of being widely used, it was not an mHealth app; this was a potential limitation, as it might have influenced respondents' perception of the impact of the questionnaire on health.) However, the sub- and total scores in other domains can be assumed to be reliable and correctly assessed by the questionnaire.

## Conclusions

The Italian version of uMARS displayed good reliability and validity. When accompanied by the "expert" MARS version, it may be used in multipurpose research/public health projects and by developers working in the sphere of digital health intervention in Italy. Furthermore, the Italian version of uMARS can provide more equal access to the evaluation of mHealth technologies from the point of view of different stakeholders (i.e. for-profit *vs* non-profit).

## Acknowledgements

## Conflicts of interest statement

The authors declare no conflict of interest.

## Authors' contributions

The authors contributed to this paper as follows: Conception and design: MS, GS, GM, SS; Administrative support: MS, GS, GM, SS; Provision of study materials or participants: BC, GL, LA, SP; Collection and assembly of data: BC, GL, LA, SP; Data analysis and interpretation: MS, SA, GM, SS; Manuscript writing: all authors; Final approval of manuscript: all authors

## References

[1] Early J, Gonzalez C, Gordon-Dseagu V, Robles-Calderon L. Use of Mobile Health (mHealth) technologies and interventions among community health workers globally: a scoping review. Health Promot Pract 2019;20:805-17. https://doi.org/10.1177/1524839919855391

[2] EUROSTAT. Key figures on enlargement countries. 2019. https://doi.org/10.2785/080458

[3] Chen X, Koskela M, Hyväkkä J. Image based information access for mobile phones. Proceedings - International Workshop on Content-Based Multimedia Indexing 2010;3:130-5. https://doi.org/10.1109/CBMI.2010.5529886

[4] Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among Australian general practitioners: survey study. JMIR Mhealth Uhealth 2019;7:e13199. https://doi.org/10.2196/13199

[5] Khachatryan K. Medical device regulation in the information age: a mobile health perspective. Jurimetrics, American Bar Association 2015;55:477-507.

[6] Modave F, Bian J, Leavitt T, Bromwell J, Harris Iii C, Vincent H. Low quality of free coaching apps with respect to the American college of sports medicine guidelines: a review of current mobile apps. JMIR Mhealth Uhealth 2015;3:e77. https://doi.org/10.2196/mhealth.4669

[7] Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001;16:671-92. https://doi.org/10.1093/her/16.6.671

[8] Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3:e27. https://doi.org/10.2196/mhealth.3422

[9] Domnich A, Arata L, Amicizia D, Signori A, Patrick B, Stoyanov S, Hides L, Gasparini R, Panatto D. Development and validation of the Italian version of the Mobile Application Rat-

ing Scale and its generalisability to apps targeting primary prevention. BMC Med Inform Decis Mak 2016;16:83. https://doi.org/10.1186/s12911-016-0323-2

[10] Makanjuola JK, Rao AR, Hale J, Bultitude M, Challacombe B, Dasgupta P. Urology apps: a review of all apps available for urologists. BJU Int 2012;110:475-7. https://doi.org/10.1111/j.1464-410X.2012.11200.x

[11] Martin Payo R, Fernandez Álvarez MM, Blanco Díaz M, Cuesta Izquierdo M, Stoyanov SR, Llaneza Suárez E. Spanish adaptation and validation of the Mobile Application Rating Scale questionnaire. Int J Med Inform 2019;129:95-9. https://doi.org/10.1016/j.ijmedinf.2019.06.005

[12] Messner EM, Terhorst Y, Barke A, Baumeister H, Stoyanov S, Hides L, Kavanagh D, Pryss R, Sander L, Probst T. The German version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR Mhealth Uhealth 2020;8:e14479. https://doi.org/10.2196/14479

[13] Tency I, Van Hecke A, Derycke J, Coorevits P. Development and validation of the Dutch version of the Mobile Application Rating Scale (MARS): a pilot study on pregnancy apps. CARE4 International Scientific Nursing and Midwifery Congress, third edition CARE4-2019 abstract book. 2019.

[14] Bardus M, Awada N, Ghandour LA, Fares EJ, Gherbal T, Al-Zanati T, Stoyanov SR. The Arabic version of the Mobile App Rating Scale: development and validation study. JMIR Mhealth Uhealth 2020;8:e16956. https://doi.org/10.2196/16956

[15] Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4:e72. https://doi.org/10.2196/mhealth.5849

[16] Panatto D, Domnich A, Gasparini R, Bonanni P, Icardi G, Amicizia D, Arata L, Carozzo S, Signori A, Bechini A, Boccalini S. An eHealth project on invasive pneumococcal disease: comprehensive evaluation of a promotional campaign. J Med Internet Res 2016;18:e316. https://doi.org/10.2196/jmir.6205

[17] George D, Mallery P. SPSS for windows step by step simple. Eleventh Edition. 2003.

[18] Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012;8:23-34. https://doi.org/10.20982/tqmp.08.1.p023

[19] Eurostat. Mobile phone penetration, 2007 and 2017 (number of subscriptions per 1,000 inhabitants) CPC19. Eurostat Statistics Explained 2019.

[20] Eurostat. Digital economy and society statistics - households and individuals. Eurostat Statistics Explained 2008;1-15.

[21] Education First. Education first: English proficiency index 2019;52.

[22] Bujang MA, Omar ED, Baharum NA. A review on sample size determination for Cronbach's alpha test: a simple guide for researchers. Malays J Med Sci 2018;25:85-99. https://doi.org/10.21315/mjms2018.25.6.9

**Correspondence:** Simone Morselli, Department of Minimally Invasive and Robotic Urologic Surgery, University of Florence, largo Brambilla 3, 50134 Florence, Italy - Tel.: +39 347 3050852 - E-mail: simone.morselli.89@gmail.com