

Applying the uMars Scale to Evaluate a Mindfulness-based mHealth App

Lakshmi Panguluri
Dept. of Computer Science)
Northern Arizona University)
Flagstaff, USA
lp755@nau.edu

Morgan Vigi-Hayes
Dept. of Computer Science)
Northern Arizona University)
Flagstaff, USA
0000-0001-9010-5380

Ashish Amresh
Dept. of Computer Science)
Northern Arizona University
Flagstaff, USA
0000-0003-3722-0720

Abstract—We present the design of a mHealth application aimed at improving mental health outcomes among young adults representing a Native American population. This study evaluates the application’s effectiveness and user-friendliness, allowing for a comprehensive understanding of its performance utilizing the uMARS (Mobile Application Rating Scale) protocol specifically designed to assess the quality of mobile health applications. Our findings indicate that the design meets both customers’ (young adults) and experts’ (mobile development practitioners) perceptions of the app. Our limitation is the lack of data collection from the population representing the Native American tribe.

Index Terms—Usability, mHealth, mindfulness

I. INTRODUCTION

Mental health is a pressing concern that requires attention and action across the globe [1]. It is important for the overall well-being of the individual and plays a significant role in shaping the ability of the individual to lead a productive and happy life. According to the World Health Organization, mental health is “a state of mental well-being that enables people to cope with the stresses of life, realize their abilities, learn and work well, and contribute to their community [2].” Mental health issues are more prevalent among Indigenous communities due to the sustained impacts of colonization [3]. The accessibility of healthcare in indigenous communities is challenging due to various factors, such as geographical isolation, insufficient cultural sensitivity, language barriers, and socio-economic issues. Addressing these challenges is crucial for advancing the accessibility and efficacy of mental healthcare for Indigenous communities. Developing culture-specific mental health interventions can gain the trust of these communities and help in the effective care of mental health [4], [5]. SUNRISE, a culturally tailored mental health intervention, is developed to cater to the mental well-being requirements of young individuals in a Native American community. Figure 1 shows the different user interaction screens from the application.

II. BACKGROUND

Native American youth—particularly those living in rural areas—are considered to be a “hard-to-reach” population in

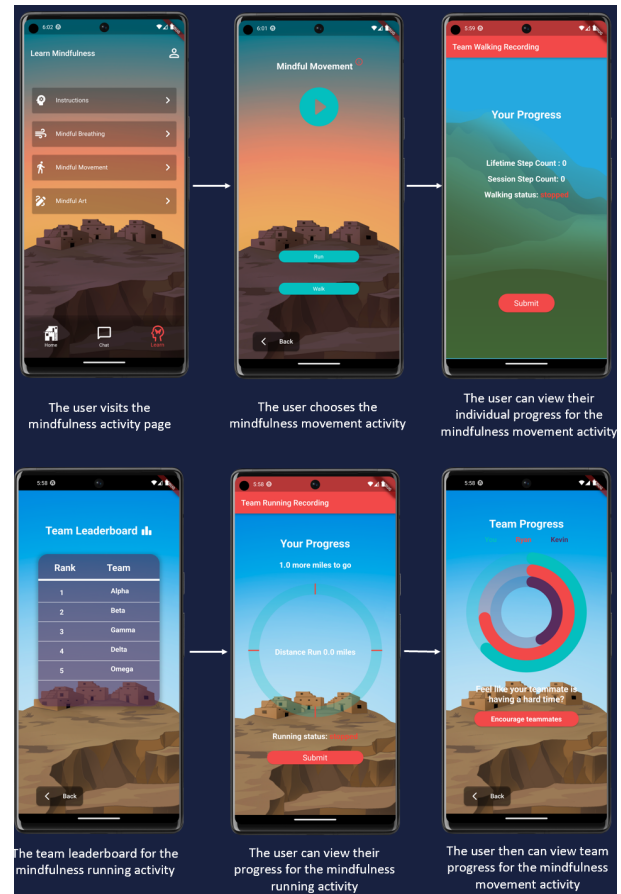


Fig. 1. User Screens from the Sunrise Mobile App

research [6], [7]. The SUNRISE app has been designed for a particular rural community where the intended users might live 1-4 of miles from the nearest next neighbor. While SUNRISE is intended to be designed to be highly culturally responsive, there is a logistical challenge of frequent engagement with target users given their rurality and the limited Internet connectivity that is available on tribal lands where they live [8]. This can make it challenging to engage in a “high-touch” user-centric design cycles that move from low-fidelity to high-fidelity prototypes through significant user input and

interaction. While our early work engaged in substantial efforts to evaluate low- and medium-fidelity prototypes with Native American youth from our community of interest, we knew that the transition to high-fidelity prototypes for pilot studies would require usability testing that did not necessitate the significant (and necessary) user input that transition from low- to medium-fidelity prototypes required. Thus, we sought to evaluate our first iteration of a high-fidelity prototype with “near-peer” users (customers) and experts. Our goal with this was to establish technical feasibility and fine-tune our initial design concept prior to deploying an initial pilot with participants from a “hard-to-reach” population.

A. SUNRISE Application

Developing mental health interventions for Native American tribal communities necessitates a culturally sensitive approach that acknowledges their deep-rooted traditions [4], [5]. At the time of our experiments, SUNRISE was a high-fidelity prototype development, and emphasized enhancing mental health and well-being in a Native American tribal community. The SUNRISE app uses a comprehensive strategy that integrates mHealth technology, educational gaming, and curricular scaffolding to promote mental wellness among adolescents. Grounded in community-based participatory research [5], the SUNRISE application incorporates activities that help teach users how to practice mindfulness skills and provide opportunities to implement those skills individually and as part of a small team.

The application operates on a team-based approach that comprises two distinct user roles: mentee and mentor. A mentee is an individual who seeks to improve their mental well-being, while a mentor is responsible for overseeing and supervising the actions of mentees. The app strongly promises anonymity for all users by allocating alias names, thereby promoting privacy and safety. The app offers a range of mindfulness activities, such as breathing exercises, running, walking, meditation, and art, all of which are geared towards improving mental wellness. While some of these activities are designed for individual use, others are intended for team-based engagement. Apart from the aforementioned activities, the app encompasses features such as messaging, a gallery, rewards, and emergency aid to augment user engagement and support.

III. RESEARCH QUESTIONS

In an effort to understand the difference between customer and expert perceptions of usability, we ask two research questions:

RQ1: How does a user’s background (experience/expertise) influence their perception of an mHealth app?

RQ2: Are there particular dimensions of usability where customers and experts tend to have more similar perceptions?

By answering these two questions, we provide insight into the impact of expertise on mobile health application evaluations, which is useful for other processes that seek to

IV. METHOD

A. Recruitment & Participants

For this study, we used purposive sampling to recruit two types of participants for user testing. The first group of participants were what we referred to as *experts*. Experts had experience developing mobile health applications or developing culturally responsive health interventions. The second group of participants were referred to as *customers*. Customers were users who were between 18-27 years old and would be representative of the age group of the target user population for the app.

After getting IRB approval for the study, we recruited expert participants by reaching out directly to individuals who were known for their experience developing mobile health apps or being involved in the development and leadership of culturally responsive health interventions. We recruited customer participants by posting physical recruitment flyers around a university campus and reaching out directly to student organizations via email. Customer participants were compensated with a \$25 Amazon gift card. This deliberate selection process prioritizes participants whose experiences and insights can significantly impact our research outcomes [9]. Overall, we recruited six customer and six expert participants ($N = 12$). None of the participants had any prior knowledge about SUNRISE. The participants except one expert were not of Native American descent.

B. Data Collection

The SUNRISE application caters to two categories of users: mentees and mentors. The mentee version is evaluated by customers, while experts assess both the mentor and mentee versions. The evaluation process involves collecting data through surveys after the participants have explored the application. The survey employs the User version of the Mobile Application Rating Scale (uMARS), a well-established and reliable tool designed to evaluate mobile health application quality. uMARS is specifically tailored for assessing mobile health applications (MHAs) and is a comprehensive tool developed through semantic analysis and literature synthesis [10]. By the recommendations of uMARS developers, participants dedicate more than 10 minutes to testing the application on their mobile devices [11]. Subsequently, they evaluate the app’s quality by completing a Google Form created using uMARS, which is a simplified user version of the Mobile Application Rating Scale (MARS). This user version consists of 20 items, condensed from the original 23. Both rating scales assess five key dimensions: engagement, functionality, aesthetics, information, and subjective quality, employing 5-point scales. The assessment of the application encompasses its influence on the user’s understanding, beliefs, and intentions about the specific health behavior. This evaluation is carried out using the perceived impact segment of the uMARS protocol [12]. The mentee version of the application was navigated by customer users who then proceeded to complete the survey. In contrast, expert users were provided with a briefing before delving into the

TABLE I
PARTICIPANT DEMOGRAPHICS.

Participant Code	Age	Gender	Race	Ethnicity
C1	22	Male	Mix of multiple races	Non-Hispanic
C2	21	Male	White	Non-Hispanic
C3	18	Female	White	Non-Hispanic
C4	18	Female	White	Non-Hispanic
C5	21	Female	Asian	Hispanic
C6	21	Female	Asian	Non-Hispanic
E1	30	Male	White	Non-Hispanic
E2	45	Male	White	Non-Hispanic
E3	33	Female	NA	NA
E4	56	Male	White	Non-Hispanic
E5	29	Male	American Indian or Alaskan Native	Non-Hispanic
E6	47	Male	Asian	Non-Hispanic

mentee version. Upon exploring the mentee version, experts were given access to the mentor version. After exploring both versions, expert users completed the survey.

V. FINDINGS

Statistical measures such as mean and standard deviation are employed in the evaluation of app features. Descriptive statistics play a pivotal role in gaining insights into the distribution of various app characteristics. Specifically, boxplots are utilized to analyze the score distributions for engagement, functionality, aesthetics, subjective quality, and perceived impact based on participants' feedback. The comparison of box plots is carried out individually for customer and expert groups. Furthermore, mean and standard deviation scores are calculated for each domain within both the customer and expert groups.

Table II displays mean scores and standard deviations for each subdomain within the domains of engagement, functionality, aesthetics, information, subjective quality, and perceived impact on the uMARS scale. In the overall ratings, Aesthetics attained the highest mean rating of 4.35, followed by Information at 3.92, Functionality at 3.71, and Engagement at 3.75. For the customer group, Aesthetics and Information obtained the highest ratings with means of 4.5, followed by Functionality at 4.21 and Engagement at 4.1. Comparatively, the expert group rated Aesthetics highest at 4, followed by Engagement at 3.4, Information at 3.33, and Functionality at 3.21. Notably, the mean total score of the app was higher for the Customer group than the Expert group. This delineates a divergence in the app's overall quality perception between the two user categories.

The following sections show the uMARS ratings of every parameter in the subdomain, along with the application's subjective quality and perceived impact. Each domain and its subdomain are explored in detail with the analysis.

1) *Engagement*: Figure 2 illustrates the overall engagement levels within the app, comparing both the customer and expert groups. The subdomains of entertainment, interest, customization, interactivity, and target group are analyzed to examine the engagement of the app. The customer group is generally skewed higher in terms of their engagement measure compared to the expert group, which is expected due to the critical lens through which the experts analyze the app.

TABLE II
UMARS RATING OF THE APP WITH A SCALE OF 1 TO 5(1-LOW AND 5-HIGH)

uMARS parameter	Overall Mean Rating (std dev.)	Customer Mean Rating(std dev.)	Expert Mean Rating(std dev.)
Engagement	3.75 (2.65)	4.1 (1.62)	3.4 (2.10)
Functionality	3.71 (2.12)	4.21 (1.20)	3.21 (1.75)
Aesthetics	4.25 (1.56)	4.5 (1.15)	4 (1.05)
Information	3.92 (2.09)	4.5 (1.17)	3.33 (1.73)
Total Score	3.91 (0.21)	4.33 (0.18)	3.49 (0.31)
Subjective Quality	3.06 (3.31)	3.25 (2.26)	2.88 (2.41)
Perceived Impact	3.81 (2.49)	4.36 (1.62)	3.25 (1.89)

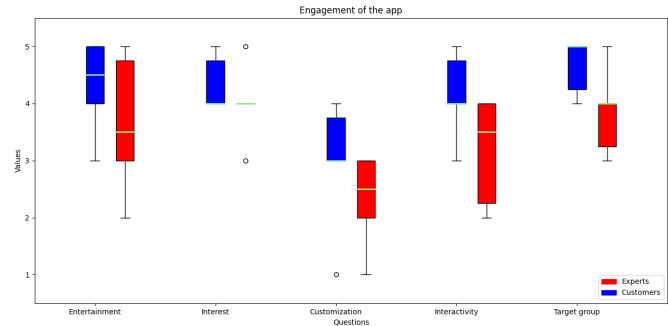


Fig. 2. Engagement of the app.

2) *Functionality*: Figure 3 displays the overall functionality ratings provided by both the expert and customer groups. The median functionality rating from the customer group surpasses that of the expert group, signaling higher satisfaction levels with the app's performance among customers. An outlier above the mean in the customer group indicates an inclination towards a higher perceived app performance. Conversely, the negative skew in the expert group's box plot suggests a slightly lower level of satisfaction with the app's performance among experts. This is also expected as the experts are asked to use both the mentee and mentor versions of the app.

3) *Aesthetics*: Figure 4 depicts the ratings of the app's aesthetics from both customers and experts. The customer group showcases notably higher satisfaction with the app's

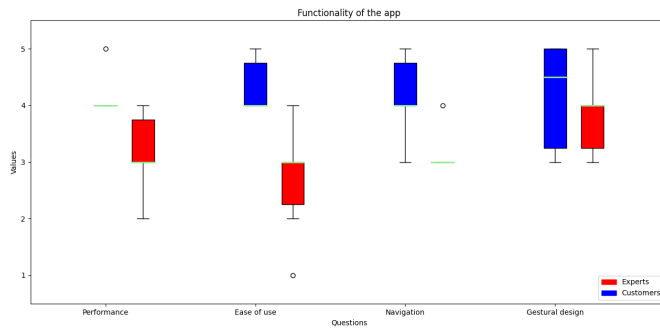


Fig. 3. Functionality of the app.

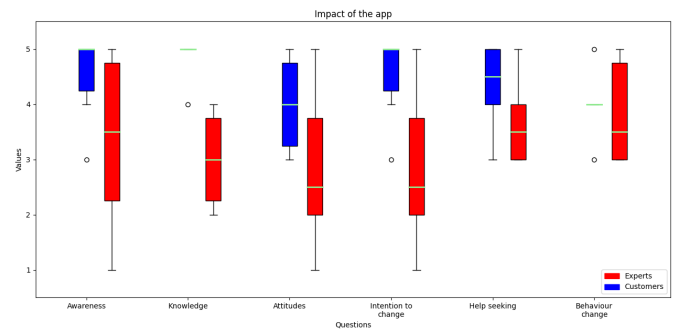


Fig. 6. Perceived Impact of the app.

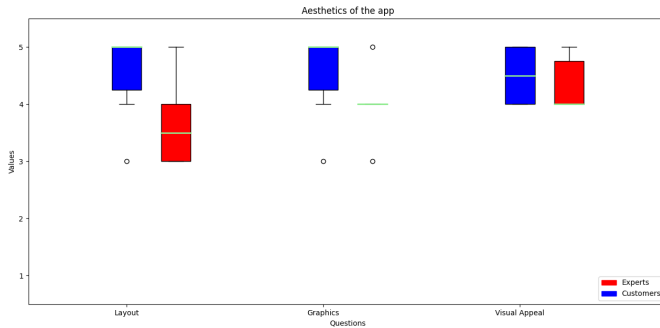


Fig. 4. Aesthetics of the app.

layout than the expert group, with a significantly higher median. Strong opinions and more positive ratings were observed within the customer group, although most inclined towards better ratings. Similar trends are observed in the appraisal of the app's graphics, with the customer group indicating a stronger preference than the expert group. However, extreme opinions regarding the graphics were noted among some experts and customers. Positive skewness among the customers represents varied opinions about the app's graphics. When considering the visual appeal, both customers and experts rated it positively. Symmetry in the box plot of customers denotes a consistent opinion about the visual appeal. At the same time, the expert group shows a positive skew, suggesting that very few rated it as exceptional.

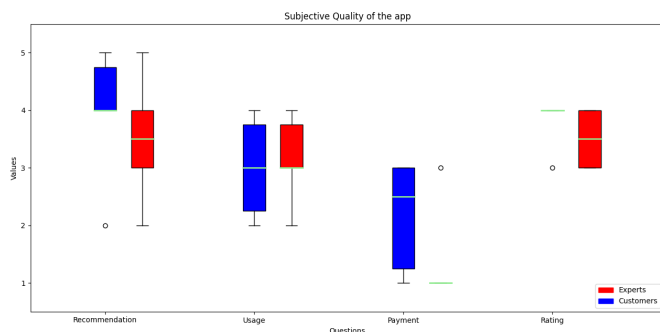


Fig. 5. Subjective Quality of the app.

4) *Subjective Quality*: The customer and expert groups display some differences in the subjective quality domain as shown in Figure 5. For recommending the app, customers lean more toward suggesting it to others than the expert group. Although experts show a wide range of balanced opinions, there is a significant disparity between the groups with no overlapping data. An outlier in the customer group represents an extreme opinion against recommending the app. However, the positive skew among customers indicates that most are inclined toward recommending the app to more people.

5) *Perceived Impact*: In the perceived impact on user knowledge, attitudes, and intentions, displayed in Figure 6, the customer group indicated a higher awareness of health behavior than the expert group. The spread in the expert group's ratings displays diverse opinions, which is evident from the widespread box. An outlier exists in the customer group, signifying a strong belief about a moderate level of awareness. Most customers strongly agreed that the app increased their knowledge of health behavior, although an outlier denotes an extreme opinion within the customer group. In contrast, the expert group shows varied opinions rather than strong agreement on increased knowledge. Experts have a more balanced range of views when rating knowledge increment.

VI. FUTURE WORK

This paper examined uMARS ratings of the SUNRISE mental health mobile application from the perspective of customers and experts with the goal of evaluating and refining a high-fidelity prototype prior to launching a series of six-week pilots with "hard-to-reach" Native American youth users. In the context of the SUNRISE application, future work will involve evaluating usability and engagement through a series of six-week pilots of the app with actual target users, which will provide an opportunity to evaluate the extent to which usability ratings from "near-peer" customers and experts approximate usability perceptions of target users.

More generally, this work contrasts the perceptions of customers and experts towards usability. In this small case study, we find that experts generally have a lower perception of usability than customers across all dimensions of the uMARS scale upon initial encounter with an application with customers

giving a mean score of 4.18 to SUNRISE compared to the mean score of 3.37 given by experts. While it makes sense that experts and customers might have different perceptions, it would be useful to validate these findings to examine whether they are generalizable to other applications and with larger N values. Moreover, future work would investigate when the gap between experts and customers begins to converge, that is, after how much experience with an application do customer users begin to have perceptions of usability that resemble those of experts? Answering these questions can help development teams as they design for hard-to-reach target users and seek to identify appropriate surrogates and calibrate the perceptions to a target user base.

ACKNOWLEDGMENT

This work was funded by NSF CNS #2224014. The authors would like to thank Vipin Verma and Shelby Hagemann for their assistance with app development.

REFERENCES

- [1] "Mental Health By the Numbers — NAMI: National Alliance on Mental Illness — nami.org," <https://www.nami.org/mhstats>, Apr 2023.
- [2] "Mental health — who.int," <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>, Apr 2022.
- [3] S. E. Nelson and K. Wilson, "The Mental Health of Indigenous Peoples in Canada: A Critical Review of Research," *Social Science & Medicine*, vol. 176, pp. 93–112, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S027795361730028X>
- [4] A. H. Alsswey, H. Al-Samarraie, F. A. El-Qirem, A. I. Alzahrani, and O. Alfarrarj, "Culture in the Design of mHealth UI: An Effort to Increase Acceptance Among Culturally Specific Groups," *The Electronic Library*, vol. 38, no. 2, pp. 257–272, 2020.
- [5] M. Vigil-Hayes, A. F. Collier, S. Hagemann, G. Castillo, K. Mikkelsen, J. Dingman, A. Muñoz, J. Luther, and A. McLaughlin, "Integrating Cultural Relevance into a Behavioral mHealth Intervention for Native American Youth," *Proceedings of the ACM on human-computer interaction*, vol. 5, no. CSCW1, pp. 1–29, 2021.
- [6] A. Shaghaghi, R. S. Bhopal, and A. Sheikh, "Approaches to recruiting 'hard-to-reach' populations into research: A review of the literature," *Health Promotion Perspectives*, vol. 1, no. 2, p. 86, 2011.
- [7] K. L. Walters, A. Stately, T. Evans-Campbell, J. M. Simoni, B. Duran, K. Schultz, and D. Guerrero, "'indigenist' collaborative research efforts in native american communities," *The Field Research Survival Guide*, pp. 146–173, 2009.
- [8] A. Bauer, D. L. Feir, and M. T. Gregg, "The tribal digital divide: Extent and explanations," *Telecommunications Policy*, vol. 46, no. 9, p. 102401, 2022.
- [9] M. Q. Patton, "Purposeful sampling and case selection: Overview of strategies and options," in *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. Sage, 2014.
- [10] Y. Terhorst, P. Philippi, L. B. Sander, D. Schultchen, S. Paganini, M. Bardus, K. Santo, J. Knitz, G. C. Machado, S. Schoeppe *et al.*, "Validation of the Mobile Application Rating Scale (MARS)," *Plos one*, vol. 15, no. 11, p. e0241480, 2020.
- [11] S. R. Stoyanov, L. Hides, D. J. Kavanagh, and H. Wilson, "Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS)," *JMIR mHealth and uHealth*, vol. 4, no. 2, p. e5849, 2016.
- [12] A. Lambrecht, N. Vuillerme, C. Raab, D. Simon, E.-M. Messner, M. Hagen, S. Bayat, A. Kleyer, T. Aubourg, G. Schett *et al.*, "Quality of a Supporting Mobile App for Rheumatic Patients: Patient-based Assessment Using the User Version of the Mobile Application Scale (uMARS)," *Frontiers in Medicine*, vol. 8, p. 715345, 2021.