EmoFrame: Prototype of a Framework to Assess Users' Emotional Responses

Suzane Santos dos Santos(✉), Erick Modesto Campos, and Kamila Rios da Hora Rodrigues

University of São Paulo, São Carlos, São Paulo, Brazil
suzanesantos@usp.br, erick.c.modesto@gmail.com, kamila.rios@icmc.usp.br

Abstract. Analyzing users' emotional aspects when interacting with computational solutions is a challenge for Computing professionals. In several situations, this kind of evaluation is the responsibility of the domain specialist. This study seeks to bring together different instruments for evaluating emotional responses in a framework named EmoFrame. It is possible to guide computer professionals in choosing the appropriate artifacts for their evaluations, depending on the solution developed and their use context. We developed a medium-fidelity prototype of the framework, and a first validation was carried out by Health and Computer specialists. It is also in the interest of this research to identify possible assessment protocols or instruments from other domains, which can be computerized with the support of these domain professionals and, later, become part of the EmoFrame.

Keywords: Computational systems evaluation · Emotional aspects · Emotional response · Framework · EmoFrame

1 Introduction

Human-Computer Interaction (HCI) is an area of research in Computer Science in which the evaluation stage is very relevant. During the evaluation, the user interface and interaction problems, not noticed in the design and development stages, are identified and corrected. This way, after a rigorous evaluation, the user has the chance to receive a safer, more effective product that does not harm they experience during the use of the product. Discussing evaluation in the context of the HCI generally leads to the concept of usability. For example, Nilsen [19] defines usability as an attribute of software quality that assesses the ease of use of user interfaces [3].

Still, in the context of evaluation and HCI, another concept intrinsically associated with usability is the concept of User Experience (UX). According to Nielsen [20], the UX covers all aspects of the end user's interaction with its services and products. More specifically, UX is about how people feel about a product and what their pleasure and satisfaction are with using it [24]. The user's emotion, in turn, is no longer just related to the system's unexpected

M. Kurosu et al. (Eds.): HCII 2022, LNCS 13516, pp. 282–301, 2022.
https://doi.org/10.1007/978-3-031-17615-9_20

response or frustration with an incomprehensible error message. The researchers now understand that a wide range of emotions plays a vital role in all tasks performed on the Computer. When interacting with computer systems, users' emotions are a fundamental aspect to help understand the user experience [29].

The recent change in the user's emotion concerning interactive systems has raised the need to understand better what emotion is and how it influences the user during the interaction. However, even though the term "emotion" is often used, there is no consensus on the concept [27]. The definition of Scherer [26] states that: "Emotion is defined as an episode of synchronized and interrelated changes in the states of all or most of the five subsystems of the organism in response to the evaluation of a stimulus event external or internal as relevant to the main concerns of the organism". Therefore, we adopted this definition in this study. The rationale behind this choice lies in the fact that this definition is one of the most comprehensive.

The project presented in this paper seeks to bring together different assessment tools, especially for assessing emotional responses, in a framework named EmoFrame. The framework has the function of: a) helping the Computer professional to find suitable tools for the target audience of the application—taking into account the particularities of the users, as well as the context of use and the requirement to be evaluated (ex.: usability, accessibility, emotional response); b) assist professionals, especially in the Health area, to apply their instruments in a computerized way and with a quick view of the results; and c) enable Computer professionals, with the support of the domain specialist, to also be able to use instruments and protocols from other areas (those possible), in order to obtain results on the effectiveness in the use of its computational solution.

In this paper, Computing instruments that can be computerized will be presented, which can facilitate their application remotely, a strategy that is especially welcome at times like the current—of social detachment due to the pandemic of COVID-19. We organized this paper into nine sections. In Related Works (Sect. 2), we discuss previous works about emotion evaluation. In the Method section (Sect. 3), we make a brief description of the techniques applied during the prototype conception. In the Selected Instruments section (Sect. 4), we detail the protocols and instruments chosen to be prototyped. In the section entitled EmoFrame Prototype (Sect. 5), we describe the screens and features of the prototyped framework. In the Evaluations and Discussion section (Sect. 6), we deal with the formal evaluation of the prototype with the project's partner specialists. Finally, we present the final considerations in the section (Sect. 7).

2 Related Works

In this section, we present some works related to the topic of interest of this research.

The study proposed by Silva et al. [32] investigates with four instruments – a set of emojis, the Self-Assessment Manikin, scroll sliders, and Semantic Emotional Space - to discover which provides information about a subjective

feeling closer to an existing emotion. The experiments conducted by the authors involved 29 volunteers taking part in four experimental rounds. A volunteer watched a movie or part of a video clip in each round and later randomly interacted with one of the instruments in a user interface. The results suggested that the scroll slider leads to more excellent proximity to the pre-classified emotions.

The work done by Xavier, Garcia and Neris [34] presents a study on the impact that elements of interfaces in computer systems have on the human emotional response. The idea was to verify how the interaction in systems with "bad" interfaces affects negative emotional responses to help build interface design. The authors chose the Ten Heuristics method [6] to evaluate users' emotions, and the target audience was older adults. The experiment consisted of carrying out a usability test of a particular system with users. Every step of user interaction with the system was filmed. The results showed that the study of emotional responses is an excellent analysis to be considered in the interface design process. This work shows the importance of evaluating users' emotional responses and how they can positively affect the quality of computational solutions.

In the study proposed by Moreira, dos Reis and Baranauskas [17], the authors developed and evaluated the TangiSAM environment, which consists of tangible artifacts designed and built to carry out assessments of affective states from the SAM, an instrument for assessing emotion often used in the area of Computing. TangiSAM includes sets of three-dimensional concrete dolls that use tangible technologies to assess affective states playfully. In this study, conducting a study in a real educational space with children and teachers is detailed to understand if TangiSAM's tangible artifacts favor a better self-assessment experience. The authors found that participants preferred TangiSAM when compared to other proposals for the representation of affective states. This study influenced our choice of an alternative form of the SAM instrument to compose the EmoFrame.

The use of emojis to assess emotional aspects occurs in the study presented by Hall, Hume and Tazzyman, [10]. The authors focus on achieving optimal responses through supporting children's judgments, using Smiley Face Likert scales as a rating scale for quantitative questions in evaluations. The paper outlines a range of studies, identifying that to achieve differentiated data and full use of rating scales by children that face positive emotions should be used within Smiley Face Likert scales. The authors used the proposed rating method (the Five Degrees of Happiness Smiley Face Likert scale) in a large-scale summative evaluation of a Serious Game. Their results highlight that the traditional Smiley Face Likert, with emotions from very happy to very unhappy, has doubtful utility as an effective method for communicating with this age group.

Our initial idea was to try to bring other works or tools that group evaluation instruments and provide a way to apply them and present results. Unfortunately, even in our best attempts, within the consulted literature, we did not find studies with this scope, so the theoretical framework is in reporting instruments for assessing emotional responses.

3 Methods

For the construction of the framework, we adopted the Participatory Design [28] methodology added to the Evolutionary Prototyping from Software Engineering [22]. In evaluating computational solutions, it is crucial to have access to the people who use the system. Participatory Design works with users to analyze claims for their current practices and then generate design ideas that address the issues raised by shared analysis. Although users generally do not have practical knowledge about the development of the application, they are very good at reacting to concrete projects that they do not like or that will not work in practice [25].

In the evolutionary prototyping technique, the developer or the development team first builds a prototype. After receiving initial customer feedback, subsequent prototypes are produced by the team, each with additional features or improvements, until the final product appears [31]. We want to emphasize that during the design/conception stage of Emoframe, we had the support of a specialist in gerontology and a specialist in psychology. They participated in workshops to support the framework's design and evaluation steps.

Initially, we conducted a study of the literature on emotional response evaluation instruments. After the study collection, there were brainstorming sessions with partner professionals; in these sessions, we discussed which instruments amongst those found in the literature (considering the case studies) we could choose. Therefore, the construction process took place through a first round in which we collected ideas and discussed the literature. Later, in a second round, we presented the prototype. Specialists validated the prototype and suggested adjustments. The next section describes the instruments chosen to compose the EmoFrame.

4 Selected Instruments

This section details the protocols and instruments that we prototyped as proof of concept for constructing the framework. The instruments selected to compose the preliminary prototype version of EmoFrame evaluate the usability of computational solutions or users' emotional responses to interactive systems. The following subsections describe such instruments. We include SUS (System Usability Scale) and SD (Semantic Differential) instruments that assess usability and UX (User eXperience) issues because we understand these issues as a requirement that leads to user satisfaction and therefore affects emotional issues.

The instruments that are part of the prototype can be applied by people other than specialists, such as psychologists or occupational therapists, which are freely distributed and have a validated translation into Brazilian Portuguese. In addition, the instruments fit the needs of the research group of the authors of this study. These needs relate mainly to two case studies, which involve both children and the elderly. The specialist also suggested an instrument that did not appear in the literature research: the WHOQOL (World Health Organization Instrument to Evaluate Quality of Life).

4.1 Geriatric Depression Scale (GDS—15)

The first instrument added to EmoFrame was the Geriatric Depression Scale (GDS) is a screening test developed, initially, by Yesavage et al. [35] and used to identify symptoms of depression in elderly. The original scale is a 30-item self-report instrument that uses a "Yes/No" response. Professionals can administer the scale with healthy adults, clinically ill adults, and those with mild to moderate cognitive impairments. The GDS scale was tested and used extensively with the elderly population. In evaluating depression in old age, the GDS scale is currently one of the most used self-reports of depression. Although specialists cannot diagnose depression exclusively based on the GDS result, they usually include its result as a part of the diagnostic evaluation due to the scale's established reliability and validity [13].

Taking into account that the GDS-30 is relatively time-consuming, an abbreviated version consisting of 15 questions (GDS-15) was developed in 1986 by Sheikh e Yesavage [30]. Among the 15 items, 10 usually indicate depression when answered positively, while the others usually indicate depression when answered negatively [9]. The scale was translated and validated for the elderly Brazilian population [1]. As mentioned above, one of the studies developed by the authors' research group is conducted with elderly people. Such studies make use of gerontology instruments to collect information about the quality of life and feelings of this population. Given this context, the GDS was incorporated into EmoFrame by: a) evaluating emotional aspects; b) attending the context of study with the elderly, being particularly important for the research group.

4.2 Profile of Mood States (POMS)

The Profile of the States of Mood is a 65-item psychological self-report instrument intended for adults aged 18 and over. POMS assesses short-term moods that are considered transient and often fluctuating [12,16]. POMS is a multidimensional Likert self-report scale, originally developed to assess the response of psychiatric patients to pharmacological and psychotherapeutic treatment. This instrument, however, quickly became applied to sport and exercise psychology, as well as to assess coping among people with chronic diseases.

The 65 items of the POMS represent six subscales that assess: tension (T), depression (D), hostility (H), fatigue (F), confusion (C), and vigor (V). A composite score—Total Mood Disturbance (TMD)—is obtained by adding five negative affect subscales and subtracting the vigor score, reflecting the total mood disturbance. POMS quickly became a trendy instrument, with adaptations for other languages. The version used in this work is the Portuguese version of the reduced version, adapted by Viana, Almeida and Santos [33]. This adapted version consists of 36 items, each of the six scales having six items. In addition, the Portuguese version (from Portugal) also features six additional items that make up the Training Misfit Scale, a complementary instrument developed by Raglin and Morgan [23], which allows assisting in the diagnosis of overtraining syndrome alerts.

Each POMS adjective is rated on a 5-point scale (0 = Never; 1 = A little; 2 = Moderately; 3 = Very; 4 = Very much). All items are quoted in the same direction, except for one item on the Tension scale and two items on the Confusion scale. In these cases, the specialist must reverse the response to the item before adding to the others. In the response instructions, we ask the users to say how they felt over a certain period. That period usually corresponds to a day or a week. In this study, we adopted the period that comprises the mood swings of the user over the last week. We obtain the POMS result in two steps: add the result of each dimension and apply the values in the TMD formula, as shown below.

$$PTH = [(T + D + H + F + C) - V] + 100 \quad (1)$$

The POMS was chosen for this work because it is a tool that assesses emotional response and can be applied to both case study audiences; we especially favored POMS because it includes a scale geared towards athletes, an area that is also our psychology collaborator's specialty.

4.3 Self-Assessment Manikin (SAM)

The Self-Assessment Manikin is an image-based questionnaire developed by Bradley and Lang [4] to measure emotional response. The questionnaire, widely used in evaluations by Computing professionals, was designed to measure three characteristics of an emotional response (pleasure, arousal and dominance), identified as central to emotion in research conducted by Lang et al. [14]. SAM can be considered free of language; that is, any individual, of any schooling, can answer it. SAM is also not limited to any culture and can be easily understood and suitable for different countries. Hayashi et al. [11] proposed an alternative form of SAM, emoti-SAM, in which they adopted different representations of the original figures. The authors created emoti-SAM due to the feedback that children gave spontaneously about the original assessment tool. According to them, most children did not like the look and colors of the original SAM. The children thought the original scale was "ugly" and did not make much sense. In response to their feedback, Hayashi et al. [11] replaced each figure in the original SAM with a corresponding emoji or emoticon—similar to those commonly used in social media and instant messaging apps. We used an adaptation of emoti-SAM in the EmoFrame.

The SAM questionnaire was chosen for this work because the version included in the framework is an adaptation aimed at children, one of the audiences that are part of the case study of interest to the research. In addition, the SAM is an instrument that is already well established in the field of Computing.

4.4 Semantic Differential (SD)

Developed by Osgood, Suci, and Tannenbaum [21], the Semantic Differential (SD) generally takes the form of a 5 or 7 point bipolar adjective scale. The

authors created this method when they realized the need to assess the affectivity and qualities of a concept and quantify the affective meaning of attitudes, opinions, perceptions, social image, personality, preferences, and interests of people [15]. Semantic scales tend to have poles and, in each pole, opposite adjectives, through which the subjects evaluate the concept, verifying the one that most expresses their feelings. One end is considered "positive" and the other end "negative", for example, stimulating and discouraging. There is a possibility of adding some questions on special interest issues, but it is usually customary to keep the questionnaire short to maximize the response rate.

SD is one instrument often used to assess people's affective perception of the objective and subjective situations faced in their daily lives. It is possible to express the concept by a word, phrase, or figure and has a psychological meaning that varies according to the group that evaluates it.

4.5 System Usability Scale (SUS)

The System Usability Scale (SUS) is a commonly used questionnaire, distributed free and reliable. The original SUS instrument was proposed by Brooke [5] and is composed of 10 statements that are scored on a 5-point agreement strength scale. The questionnaire score results in a usability score in the range 0–100. A positive feature of SUS is that it provides an exclusive reference score for participants' opinions on the usability of a product. The ease of administration and scoring of SUS makes it a popular choice among usability professionals. In addition to being a popular choice for online usability research, SUS can be used as a subjective follow-up measure after testing the usability of functional systems as a pre-and post-test component [2,7].

We choose an adaptation of the original SUS to compose the prototype of the EmoFrame. The adapted version is composed of 28 items that comprise the ten original items.

4.6 The World Health Organization Instrument to Evaluate Quality of Life (WHOQOL-BREF)

The WHOQOL Group developed the WHOQOL-100 quality of life assessment with fifteen international field centers simultaneously to develop a quality of life assessment cross-culturally applicable.

The WHOQOL-100 allows a detailed assessment of each facet related to the quality of life. In some instances, however, the WHOQOL-100 may be too long for practical use. Therefore, the WHOQOL-BREF was developed to provide a summary assessment of the quality of life that analyzes domain level profiles, using data from the WHOQOL-100 pilot assessment. An item from each of the 24 facets of the WHOQOL-100 was included to provide a broad and comprehensive assessment. In addition, two items from the General quality of life and General health facet were included (questions 1 and 2) [8].

We choose the WHOQOL for this work at the suggestion of a specialist in gerontology. WHOQOL is an instrument validated by the WHO and evaluates

very relevant aspects for the population in general. Our interest is to know how the elderly population perceives their quality of life and health.

The next section describes how the framework containing the aforementioned instruments was developed.

5 EmoFrame Prototype

After analyzing and choosing the instruments to be adopted in the EmoFrame, which we will use in two case studies, we implemented an interactive prototype (medium-high fidelity) from EmoFrame. We emphasize that the term framework is used in the context of this work in the broadest sense as a structure composed of mechanisms, artifacts, and systems used in planning and decision-making regarding software evaluation. The prototyped interface of EmoFrame, available in the images below, represents the evaluated version to specialists to assess with their populations of interest. The specialists, in general, will have access to the instruments and be able to register users and could consult the results of the evaluations conducted by them. The user will access the system when registered by a specialist and have access only to the tools and not to the results.

The EmoFrame provides accessibility features, such as: increase the font, decrease the font, and contrast. To register, the specialist must provide the following information: Name; Social Name; Specialty; Phone Number; Gender; Date of birth; E-mail; and Password. After completing the registration, the specialist can enter the system by providing an e-mail and password.

5.1 Used Instruments and Instructions

Upon entering the system, the specialist has access to three main pages: tools, registering users, and results. The Fig. 1 illustrates the tool screen.

This tab contains the six instruments selected to compose the framework, until now. All instruments have an initial sentence, a kind of instruction to answer the test. In addition, they also have a button called "instructions" that contains examples of how to answer the scales.

5.2 User Registration on EmoFrame

In the second tab, the specialist can register new users to have access to the tools. The register is essential so that the different data are collected so that possible correlations between the data obtained through the questionnaires and sociodemographic data, for example, can be traced. In addition, with the user registered on the system, it is possible to store, in a safe and adequately anonymized manner—if necessary, the results of the instruments and specialists will be able to access this data whenever necessary. The requested data are as follows: Name; Social Name; Nationality; Naturalness; Address; Telephone; Gender; Marital Status; Education; Individual Monthly Income; Monthly Family Income; Date of birth; and Email.

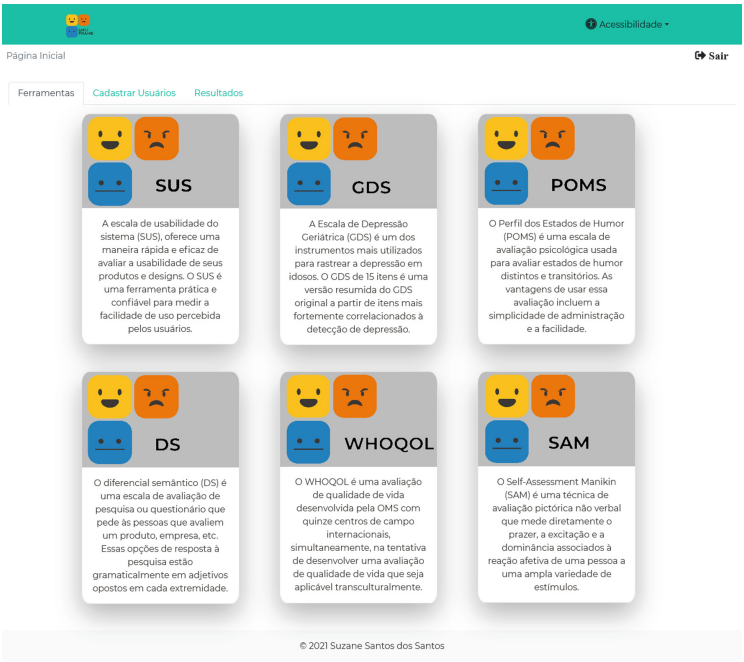

Fig. 1. Instruments—EmoFrame.

5.3 EmoFrame Results

The results are available to the specialists who applied the instruments. A specialist who applied SAM, for example, to 100 users, will only have access to the data of those users. Access control is a crucial part of protecting this sensitive data. The results can be filtered by the user's name or by the six available instruments. The following subsections describe how the results are made available for each instrument.

SUS Results. We divided the SUS result into three types of visualizations. The first is a score combined with a reference table, described in Table 1. The second view is a horizontal bar chart that illustrates all the questions and answers given by the user.The last response visualization is a "curve" that illustrates the response variation. In addition, at the bottom of the screen, there is a button that, when clicked, illustrates how we calculate the SUS score.

GDS Results. The GDS result has a score and a reference table, in addition to the button on how to calculate. The reference we adopted is shown in Table 2.

Table 1. SUS score reference table.

Score	Usability Status
Less than 60	Unacceptable
60–70	Ok.
70–80	Good.
80–90	Excellent.
Greater than 90	Best possible usability

Table 2. GDS reference table.

Score	Result
1–5	Low risk for depression indication
6–10	Moderate risk for depression indication
11–15	Severe risk for depression indication

POMS Results. The result consists of two different scores, a TMD result and a result referring to the training misfit scale. We show the results references of Total Mood Disturbance (TMD) in Table 3.

Table 3. POMS reference table.

Score	Result
High TMD score	Indicate worsening of mood
Low TMD score	Indicate improvement of mood

To know the result of the training misfit scale, add the results of the questions related to this domain to the PTH. A high value indicates overtraining, and a low value implies that the exercise practitioner is healthy.

SD Results. We illustrate the SD result using a chart with two axes representing opposite poles. We plot the user responses in the form of a line that demonstrates the user's trend. In addition, the SD results page also contains a table with responses, which range from -3 to 3.

WHOQOL Results. The WHOQOL result is divided between the scores for four domains and the answers for two general questions. Each domain has an individual score, and the domain scores are scaled positively. That is, higher scores indicate higher quality of life. The results page contains a table showing each score for each domain, a vertical bar chart, and a table showing the score for the two general questions.

SAM Results. SAM evaluates three domains, and its result is given in the form of a table, where each row represents one of the domains. The table is graded by colors and each color indicates the tendencies of the responses: red—negative; yellow—neutral; and green—positive. Examples of results from each of the aforementioned instruments can be viewed in the project's repository[1].

5.4 Final Considerations About the Prototype

The Computing area is broad and plural; that is, it covers many areas and, consequently, the most diverse users. Therefore, the motivation to build EmoFrame arose from finding appropriate instruments to evaluate specific solutions that meet the most diverse needs scenarios. Furthermore, because we believe that evaluating the quality/efficiency of solutions is a difficulty faced by other areas, we seek to make EmoFrame a valuable tool for many professionals. Hence, we intend that specialists access the framework and filter an instrument that suits their solution and meets the target audience's demands. In addition, we expect that, even if these specialists have never seen the instrument, they will use it correctly, following the instructions and guidelines that EmoFrame may offer.

The results pages are essential so that specialists from other research areas can understand users' opinions and feelings when interacting with solutions or interventions proposed by them. The results of SUS and SD show an overview of the usability and quality of solutions. From the users' answers, the tester can know the strengths and weaknesses of the evaluated solution. With the GDS result it is possible to refer the user to a more rigorous and detailed assessment if the user presents a high indication of depression, so the instrument can be considered a powerful way of screening for some disorders. The WHOQOL result is an essential piece of data to understand the users' perception of quality of life and health. It is an instrument that assesses, above all, the user's understanding of situations such as housing and support systems. Although simple, the SAM result tells us if a user felt good when using a particular solution, if users feel in control (safe, for example) and if what was proposed by the developers can motivate them.

6 Evaluation and Discussion

In the first evaluation of the framework, the specialists (psychologist, gerontologist, and HCI specialists) freely evaluated all EmoFrame screens. We did not pre-set any script or scenario. During the interaction with the prototype, they asked several questions and suggestions, such as increasing the font size, changing the initial sentences of some instruments, adding or removing buttons, for example. All EmoFrame screens have changed after this first assessment, but in general, the specialists approved most of the features of EmoFrame, considered an interface "clean" and very similar to the instruments applied on paper.

[1] https://drive.google.com/file/d/1wGMyyJlItWHtI2ThUReVwibRwdQ4Xozn/view?usp=sharing.

After the first evaluation, we conduct a second evaluation (user test) with the specialists who participated in the previous evaluation (gerontologist and psychologist). We also invited three other specialists in HCI, who are also part of the research group. To conduct the validation, we set up three different fictional scenarios: A, B, and C. For the three scenarios, we ask specialists to carried out a series of tasks, namely: (1) To access the EmoFrame; (2) To carry out the registration as a specialist; (3) To register a user; (4) To evaluate the tools of the scenario to which they were assigned; (5) To open the results page of the instruments evaluated by them; (6) To answer the SUS questionnaire, evaluating the EmoFrame and (7) Log out.

In scenario A, the personas are elderly people who participate in a digital literacy course. In this context, users learned to use various applications for smartphones and tablets, such as Instagram® and YouTube®. At the end of the course, it is interesting for the researchers responsible for students' course feedback on their experience. In this scenario, the focus is on evaluating the students' experience during the course, for which we used two instruments. The first is the GDS, and the second is WHOQOL. As the public in this scenario are elderly people, the gerontologist evaluated the instruments of this scenario.

In scenario B, the personas are high school students practicing sports and sedentary students. In the suggested context, the school board asked the institution's psychologist to investigate the students' emotional state and compare the results of active students of some sport (who were part of a team) with students who do not practice any sport actively. The instruments used were POMS and SD. The POMS is a scale for assessing mood states and also includes the Training Misfit Scale (TMS). To report how they felt when answering POMS, the students answered the SD. The specialist responsible for evaluating this scenario was the psychologist.

In scenario C, the specialists used EmoFrame to evaluate an educational game. The game has three phases. At each stage played, the specialists responded to the SAM to say how they felt when interacting with the computational solution. When completing this step, the specialists should carry out the tasks listed above, evaluating the SAM and SUS instruments. We have assigned the three HCI specialists to this scenario.

6.1 Scenario Discussions

When starting the evaluation of Scenario A, the specialist noticed that when she entered the system after registering, her username did not appear anywhere on the screen, leaving her confused and uncertain about completing the registration. Another observation made by the gerontologist concerns the initial WHOQOL sentence. The specialist stressed the importance of informing the user that the World Health Organization validates the instrument. One of the main contributions made in this scenario involves the date of the test. The specialist said this topic: *"It is important that I have the results, the name, the age and the date of application of the test. With the date of application, you can follow the evolution of the users if they are under treatment."*.

Regarding scenario B, the psychologist made several suggestions. One of them was to include an option in all questionnaires that says: "I do not know/I do not want to answer" according to the specialist, users do not have to answer what they do not want. Regarding POMS, the psychologist believes that it is better to remove the numbers from the POMS questions and change the order of the items if the intention is to do a "pre and post test". In addition to these specific suggestions, the specialist stated that it would be positive to add a field to insert comments on the questionnaires in the form of text, video, or audio. Regarding the user registration, the suggestions were to add a field to inform who is responding to the user registration (responsible or the individual himself), add the field "race" to the questionnaire, and put the options in alphabetical order. He also suggests dividing the fields "The first name" and "The last name", adding an option "I do not know/I prefer not to say" in the gender question. Regarding the SD instruction, the psychologist said: *"The instructions are perfect, it is important to define what is between the extremes. The example with the 3 points is the recommended one".*

In scenario C, HCI specialists suggested adding feedback whenever the user sends a response or fills out a form. In addition, they stressed the need to include a field for reporting errors. One of the specialists made the following comment: *"I really liked it. I found the interface very well done. The emojis used, [...] were very well chosen. As much as the person doesn't read the description, you can understand what each one represents. The texts are clear and objective."*

6.2 SUS Evaluation

After carrying out the requested tasks, the specialists answered the SUS instrument about the EmoFrame interface. We categorize SUS questions according to Nilsen's usability heuristics [18].

Match Between System and the Real World. According to the definition of this heuristic, the design must follow real world conventions, making the information appear in a natural and logical order. The SUS questions that we consider to fall into this category are listed below:

- Question 8: *"Navigating the application's menus and screens was easy."*;
- Question 12: *"It is easy to remember how to do things in this application."*;
- Question 23: *"The terminology used in the button texts was easy to understand. ".*

Figure 2 illustrates the responses of the 5 specialists who evaluate the framework. The positive responses regarding this heuristic infer that terms, concepts, icons and images seem perfectly clear to users.

User Control and Freedom. We consider that the questions below are related to this heuristic:

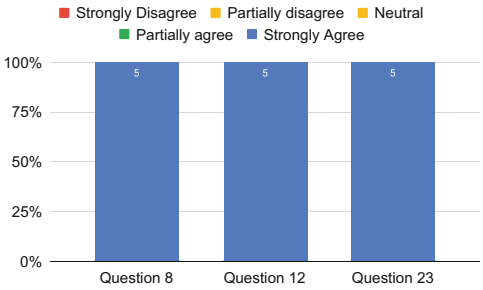

Fig. 2. Match between system and the real world.

- Question 4: *"I felt in charge using this app."*;
- Question 7: *"It is easy to do what I want using this application."*;
- Question 25: *"I felt comfortable using this app."*.

Based on the users' responses, shown in Fig. 3, we can consider that the framework promotes a sense of freedom and confidence to the specialists. Only one specialist was neutral in this category when it came to one of the questions. The specialist would like more tools available in the framework.

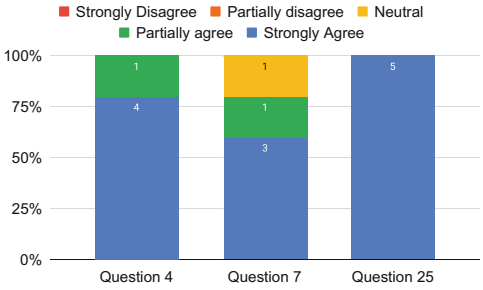

Fig. 3. User control and freedom.

Consistency and Standards. Conceptually, according to this heuristic, users should not ask themselves if different words, situations, or actions mean the same thing. That is, the system follows a pattern. The questions regarding this concept are listed below:

- Question 11: *"I found the app consistent. For example, all functions can be performed similarly."*;
- Question 20: *"The symbols and icons are clear and intuitive."*;
- Question 26: *"The application behaved as I expected."*;

– Question 28: *"I found that the various functions of the application are well integrated."*.

The 5 specialists agree that the tool is consistent and follows a well-established pattern, as shown in Fig. 4.

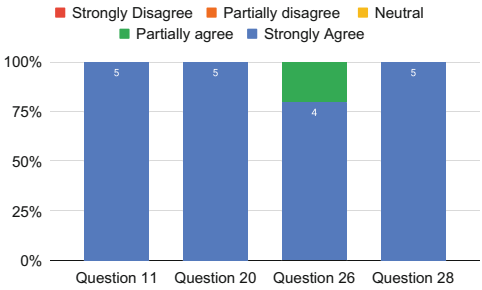

Fig. 4. Consistency and standards.

Help Users Recognize, Diagnose, and Recover from Errors. Two questions present in the SUS related to this heuristic are listed below. According to Nilsen, the error messages must be expressed in simple language, accurately indicate the problem, and suggest a solution constructively.

– Question 2: *"When I make a mistake, it is easy to correct it."*;
– Question 3: *"Error messages help to correct problems."*.

The users' responses show a flaw in the framework. According to the specialists, they are not faced with any error message and no field to report these errors. We consider this as a usability problem. The responses of the 5 specialists can be seen in Fig. 5.

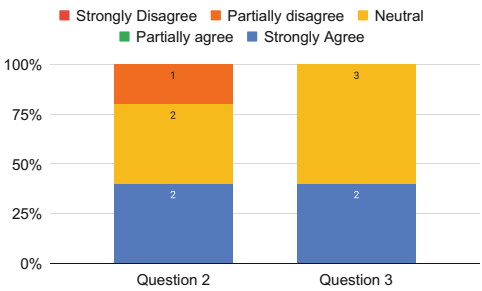

Fig. 5. Help users recognize, diagnose, and recover from errors.

Recognition Rather Than Recall. This heuristic says that the user should not remember information from one part of the interface to another. That is, the information needed to use the design must be visible or easily retrievable when necessary. The questions listed below refer to this Nielsen's heuristic:

- Question 6: *"It was easy to learn how to use this app."*;
- Question 14: *"The organization of menus and action commands (such as buttons and links) is logical, allowing you to find them easily on the screen."*;
- Question 17: *"The application provides all the information necessary to complete the tasks clearly and understandably."*.

The specialists' positive responses, shown in Fig. 6, to the EmoFrame interface suggest that it promotes recognizing actions and reduces the amount of cognitive effort required from users to carry out tasks within the framework.

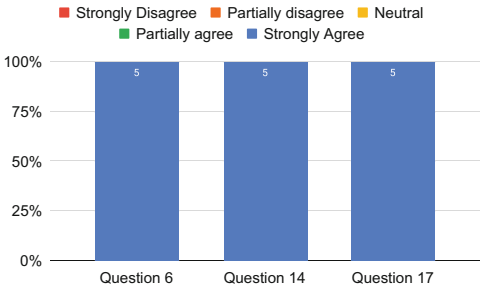

Fig. 6. Recognition rather than recall.

The questions below are also related to this heuristic. However, they make negative statements about the framework. Therefore, the adverse responses are positive about EmoFrame, as we show in Fig. 7.

- Question 18: *"I found the app very complicated to use."*;
- Question 19: *"I needed to learn many things to use this application."*;
- Question 22: *"I found the application unnecessarily complex. I had to remember, research or think hard to complete the tasks."*;
- Question 24: *"I would need support from a person to use this app."*.

Flexibility and Efficiency of Use. This heuristic concerns efficiency when executing actions within the system; flexibility implies that the different users can execute the many processes in different ways to choose the method that works for them. The related questions are:

- Question 1: *"I found it easy to enter data into this application."*;
- Question 5: *"I thought the time it took to complete the tasks was adequate."*;

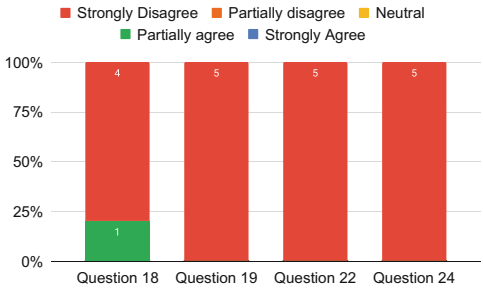

Fig. 7. Recognition rather than recall (reverse scale).

– Question 9: *"The application meets my needs."*;
– Question 10: *"I would recommend this app to others."*;
– Question 13: *"I would use this app frequently."*;
– Question 16: *"I enjoyed using this app."*.

We show the specialists' responses in Fig. 8.

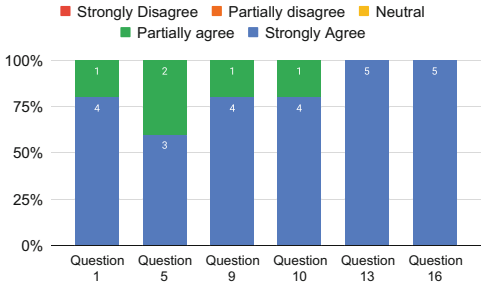

Fig. 8. Flexibility and efficiency of use.

Aesthetic and Minimalist Design. Nielsen says that the interfaces should not contain irrelevant or rarely needed information. Each extra unit of information in an interface competes with the relevant information units and decreases their relative visibility. The two SUS questions that we consider related to these aspects are:

– Question 15: *"The app's interface design is attractive."*;
– Question 21: *"I found the texts easy to read."*.

We show the specialists' responses in Fig. 9. According to the answers, we can infer that the design pleased the users or that at least it did not bother them to the point of impairing the interaction.

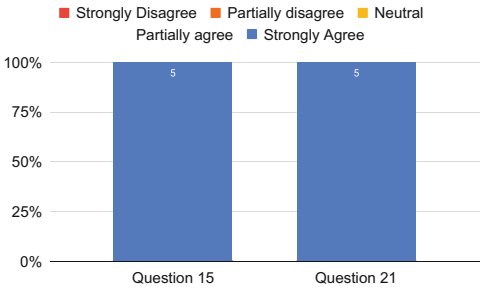

Fig. 9. Aesthetic and minimalist design.

7 Final Remarks

We presented a medium-high fidelity prototype of the EmoFrame. The technologies we used to develop the prototype were HTML, CSS, and JavaScript. We concluded during this study that it is also necessary to include guidelines to compose the framework and a set of instruments. As a contribution, we hope to offer the EmoFrame as an artifact, together with the synthesized instruments.

A work in progress is the development of the EmoFrame in high fidelity. It is in the interest of the research group to carry out new studies also in the tool's interface, as well as to use it in the context of research with the populations of interest of the group's specialists. In the future, it is expected that the specialist can receive recommendations for instruments to be used in their studies, based on input data offered by these specialists, such as the desire to assess emotional responses with children, the elderly, carriers disability, among other audiences, for example, and also for different types of emotional responses.

Acknowledgments. The authors would like to thank the financial support from Brazilian agency: CAPES and also to thank the participants in the evaluation.

References

1. Almeida, O.P., Almeida, S.A.: Confiabilidade da versão brasileira da Escala de Depressão em Geriatria (GDS) versão reduzida. Arquivos de Neuro-Psiquiatria. **57**, 421–426 (1999). http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0004-282X1999000300013&nrm=iso
2. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the system usability scale. Int. J. Human Comput. Interact. **24**(6), 574–594 (2008). https://doi.org/10.1080/10447310802205776
3. Barbosa, S., Silva, B.: Interação Humano-Computador. Elsevier Brasil (2010). https://books.google.com.br/books?id=qk0skwr_cewC
4. Bradley, M.M., Lang, P.J.: Measuring emotion: the self-assessment manikin and the semantic differential. J. Behav. Ther. Exp. Psych. **25**(1), 49–59 (1994)
5. Brooke, J.: SUS: a quick and dirty usability scale. Usabil. Eval. Ind. **189**, 4–7 (1995)

6. De Lera, E., Garreta-Domingo, M.: Ten emotion heuristics: guidelines for assessing the user's affective dimension easily and cost-effectively. In: Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK, vol. 21, pp. 1–4 (2007)
7. Finstad, K.: The system usability scale and non-native English speakers. J. Usabil. Stud. **1**(4), 185–188 (2006)
8. Fleck, M.P., et al.: Aplicação da versão em português do instrumento abreviado de avaliação da qualidade de vida "WHOQOL-bref". Revista de Saúde Pública **34**, 178–183 (2000). http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102000000200012&nrm=iso
9. Greenberg, S.A.: The geriatric depression scale (GDS). Best Pract. Nurs. Care Older Adults **4**(1), 1–2 (2012)
10. Hall, L., Hume, C., Tazzyman, S.: Five degrees of happiness: Effective smiley face Likert scales for evaluating with children, pp. 311–321, June 2016. https://doi.org/10.1145/2930674.2930719
11. Hayashi, E.C.S., Posada, J.E.G., Maike, V.R.M.L., Baranauskas, M.C.C.: Exploring new formats of the self-assessment manikin in the design with children. In: IHC 2016, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/3033701.3033728
12. Heuchert, J.P., McNair, D.M.: Profile of mood states 2nd edition™ (2012)
13. Lach, H., Chang, Y.P., Edwards, D.: Can older adults with dementia accurately report depression using brief forms? J. Gerontol. Nurs. **36**, 30–37 (2010). https://doi.org/10.3928/00989134-20100303-01
14. Lang, P.J., Greenwald, M.K., Bradley, M.M., Hamm, A.O.: Looking at pictures: affective, facial, visceral, and behavioral reactions. Psychophysiology **30**(3), 261–273 (1993)
15. Lopes, J.L., Nogueira-Martins, L.A., Andrade, A.L., Barros, A.L.B.L.: Escala de diferencial semmântico para avaliação da percepção de pacientes hospitalizados frente ao banho. Acta Paulista de Enfermagem. **24**, 815–820 (2011). http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-21002011000600015&nrm=iso
16. McNair, D.M., et al.: Manual profile of mood states (1971)
17. Moreira, E., dos Reis, J., Baranauskas, M.C.: Artefatos tangíveis e a avaliação de estados afetivos por crianças. Revista Brasileira de Informática na Educação. **27**, 58 (2019). https://doi.org/10.5753/rbie.2019.27.01.58
18. Nielsen, J.: 10 usability heuristics for user interface design. https://www.nngroup.com/articles/ten-usability-heuristics/ (1994). acesso em 02 Fev 2021
19. Nielsen, J.: Usability 101: Introduction to usability. https://www.nngroup.com/articles/usability-101-introduction-to-usability/ (2012). acesso em 02 Fev. 2021
20. Norman, D., Nielsen, J.: The definition of user experience (UX). https://www.nngroup.com/articles/definition-user-experience/ (2006). acesso em 02 Fev. 2021
21. Osgood, C.E., Suci, G.J., Tannenbaum, P.H.: The Measurement of Meaning. No. 47, University of Illinois press (1957)
22. Pressman, R., Maxim, B.: Engenharia de Software-8ª Edição. McGraw Hill, Brazil (2016)
23. Raglin, J., Morgan, W.: Development of a scale to measure training-induced distress. Med. Sci. Sports Exer. **21**, S50 (1980). https://doi.org/10.1249/00005768-198004001-00299
24. Rogers, Y., Sharp, H., Preece, J.: Interaction Design: Beyond Human-computer Interaction. John Wiley & Sons, Hoboken (2011)
25. Rosson, M.B., Carroll, J.M.: Usability Engineering: Scenario-Based Development of Human-computer Interaction. Morgan Kaufmann, Burlington (2002)

26. Scherer, K.R.: Appraisal considered as a process of multilevel sequential checking. Apprais. Process. Emot. Theory Methods Res. **92**(120), 57 (2001)
27. Scherer, K.R.: What are emotions? and how can they be measured? Soc. Sci. Inf. **44**(4), 695–729 (2005). https://doi.org/10.1177/0539018405058216
28. Schuler, D., Namioka, A.: Participatory Design: Principles and Practices. CRC Press, Boca Raton (1993)
29. Sears, A., Jacko, J.A.: Human-computer Interaction Fundamentals. CRC Press, Boca Raton (2009)
30. Sheikh, J.I., Yesavage, J.A.: Geriatric depression scale (GDS): recent evidence and development of a shorter version. Clin. Gerontol. J. Aging Mental Health. (1986)
31. Sherrell, L.: Evolutionary Prototyping. In: Runehov, A.L.C., Oviedo, L. (eds.) Encyclopedia of Sciences and Religions, pp. 803–803. Springer, Netherlands, Dordrecht (2013). https://doi.org/10.1007/978-1-4020-8265-8_201039
32. Silva, L.G.Z., Guimarães, P.D., de Souza Gomes, L.O., de Almeida Neris, V.P.: A comparative study of users' subjective feeling collection instruments. In: Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems, pp. 1–10 (2020)
33. Viana, M.F., Almeida, P.L.d., Santos, R.C.: Adaptação portuguesa da versão reduzida do Perfil de Estados de Humor: POMS. Análise Psicológica. **19**, 77–92 (2001). http://www.scielo.mec.pt/scielo.php?script=sci_arttext&pid=S0870-82312001000100008&nrm=iso
34. Xavier, R., Garcia, F., Neris, V.: Decisões de design de interfaces ruins e o impacto delas na interação: um estudo preliminar considerando o estado emocional de idosos, pp. 127–136 (2012)
35. Yesavage, J.A., et al.: Development and validation of a geriatric depression screening scale: a preliminary report. J. Psych. Res. **17**(1), 37–49 (1982)