Developing an Emoji-based User Experience Questionnaire: UEQ-Emoji

Ashley Colley
University of Lapland
Rovaniemi, Finland
ashley.colley@ulapland.fi

Sven Mayer
LMU Munich
Munich, Germany
info@sven-mayer.com

Jonna Häkkilä
University of Lapland
Rovaniemi, Finland
jonna.hakkila@ulapland.fi

Figure 1: The 11-item emoji-based UEQ showing the mapping to the UEQ subscales. Note: The novelty subscale is not supported due to semantic limitations in the emoji language.

ABSTRACT

Emojis are increasingly used in tools to evaluate user opinions, typically through happy-to-sad rating scales. We extended this concept by developing an emoji-based User Experience Questionnaire (UEQ), a well-established UX evaluation tool. Our design process involved selecting emojis corresponding to UEQ adjectives and creating bipolar emoji scales. The final tool covered 11 scales but could not represent the *novelty* subscale due to limitations of the emoji language. A comparative user study showed statistically similar results for 7 out of 11 scales between the emoji-based and word-based UEQs. Although the word-based version was preferred, our findings suggest that emoji-based scales can be a viable alternative for UX evaluation. Our work provides new insights into the potential and limitations of using emojis in UX research. Supplementary materials include a ready-to-use UEQ-Emoji form and an analysis tool.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

MUM '23, December 03–06, 2023, Vienna, Austria

ACM ISBN 979-8-4007-0921-0/23/12.
https://doi.org/10.1145/3626705.3627767

KEYWORDS

Emojis, User Experience (UX), UX Evaluation Tools, User Experience Questionnaire (UEQ), Bipolar Differential Rating Scales, Survey Design, User Study

ACM Reference Format:
Ashley Colley, Sven Mayer, and Jonna Häkkilä. 2023. Developing an Emoji-based User Experience Questionnaire: UEQ-Emoji. In *International Conference on Mobile and Ubiquitous Multimedia (MUM '23), December 03–06, 2023, Vienna, Austria.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3626705.3627767

1 INTRODUCTION

Emojis have become ubiquitous in digital communication, with over 3,521 emoji currently recognized by the Unicode Consortium. For example, the smiling-face-with-smiling-eyes emoji (Unicode: U+1F60A name: smiling face with smiling eyes) is widely used to convey happiness, friendliness, or gratitude. Emojis can express emotions in a concise and visually engaging manner [7]. When included as part of text-based messaging, emojis can add tone of voice or facial expressions and form a crucial element in correctly understanding the meaning of a message [39]. A considerable body of research, including Lu et al. [24], has delved into the interpretation of emojis by individuals, underscoring that emojis do not constitute a universal language. Emojis have also become increasingly popular as a response method, particularly online surveys. Popular survey tools such as Google Forms, SurveyMonkey, and Qualtrics include the option to use emojis as a response method, allowing participants to choose from a pre-selected list of emojis when answering questions.

As well as selecting individual emojis, happy-to-sad rating scales using emojis have become a ubiquitous form of opinion data collection, e.g., physical smiley feedback terminals are a common sight in shops and public service facilities and present a scale using four emoji buttons [16]. Where such bipolar differential rating scales, i.e., with reference terms at both ends to measure both the positive and negative sides of the response spectrum, have been implemented using emojis, they almost exclusively also present variations on a happy-to-sad scale [3, 4]. Commonly used UX evaluation tools that utilize a bipolar differential rating scales format include Attrakdiff and the User Experience Questionnaire (UEQ) [23, 33]. In the UEQ, respondents are asked to rate their overall experience using bipolar adjective pairs, such as *annoying–enjoyable* or *complicated–easy*. Each pair forms a bipolar scale, and the respondent's rating on each scale contributes to the overall evaluation of the user experience [23, 33].

It is interesting to note that, similarly to the well-discussed differences in understanding the meaning of emojis, even words, the basic units of language, are subject to varying interpretations. For instance, seemingly straightforward words like *penguin* can be understood differently [25]. This suggests that words used in existing user experience evaluation scales, such as *conventional* or *unpredictable*, will likely be interpreted differently. Such terms may appear straightforward but are complex and nuanced, open to various interpretations.

Given the propensity of emojis to communicate feelings, we explore the development of a UX evaluation tool using emojis as the endpoints of bipolar differential rating scales. To ground the design, we based our development on the well-established UEQ tool. Through an iterative process, see Figure 3, we first established a set of emojis that are understood to represent the individual adjectives in the UEQ. We then formed scales using bipolar emoji pairs and collated these to produce an emoji version of the UEQ. We report that the *novelty* subscale of the UEQ is not easily represented by emojis. Based on a user study comparing word and emoji-based UEQs, we report statistically similar results in 7 of 11 rating scales. While there were no quantifiable differences between participants' performance in the two cases, the word-based tool was preferred. We contribute by:

(1) Exploring emoji-based bipolar differential rating scales beyond the ubiquitous happy-to-sad scale. For example, using emojis to represent exciting-to-boring.
(2) Developing a UX evaluation tool comprised of a set of emoji-based bipolar differential scales to provide data on several facets of the user experience.
(3) Validating the developed emoji-based UEQ through a comparative user study.
(4) We provide a ready-to-use UEQ-Emoji form and analysis tool as supplementary materials to this paper.

2 RELATED WORK

To position our work, we provide background by referencing relevant works in UX measurement tools and emoji use in communication. We then discuss prior works that have utilized emojis as user feedback tools and highlight our contribution.

2.1 UX Measurement Tools

In the realm of UX measurement, Hassenzahl and Tractinsky [17] laid the foundational framework, emphasizing that both practical (usability) and hedonic dimensions contribute to the overall UX. UX can be measured using qualitative methods such as interviews, text-based questionnaires, and the think-aloud method when using a product or service [38]. For quantitative UX research, rating scales and Likert scales are the basic tools [35]. In Likert scales, respondents express their agreement with statements on a scale, e.g., ranging from 1 (totally disagree) to 7 (fully agree). Building on these basic rating scales, several widely used UX evaluation instruments have emerged, e.g., the System Usability Scale (SUS), the NASA Task Load Index (NASA-TLX), AttrakDiff, and the User Experience Questionnaire (UEQ) [11]. While the SUS and NASA-TLX are tailored to focus on the usability of products and services, AttrakDiff and the UEQ are designed to capture a holistic view of the UX, encompassing both usability and hedonic dimensions of the experience [11].

Rating scales may be either unipolar or bipolar. Unipolar rating scales present a continuum, such as the degree of agreement, from zero to a defined maximum. In contrast, Bipolar scales present a continuum between opposites (positive and negative), such as the degree of agreement and disagreement [18]. Of the previously mentioned UX evaluation tools, only NASA-TLX utilizes a unipolar format, e.g., rating mental demand from very low to very high. The SUS uses a bipolar scale with ratings from strongly disagree to strongly agree, and both AttrakDiff and the UEQ use opposing adjectives as the scale endpoints. DeCastellarnau [10] presents a useful literature review of response scale characteristics, noting that while bipolar scales can measure neutrality, the direction, and the intensity of opinion, unipolar scales only measure the intensity. Prior work has reported that bipolar rating scales tend to have a positive bias, as respondents are reluctant to choose negative responses, while unipolar scales have a center bias [10, 18].

2.2 Emojis in Communication

Emojis have emerged as part of the online messaging culture to enrich textual communication, and their use has a role in counteracting the lack of non-verbal cues in online messaging channels [40]. Early works on emoticons and emojis report how smiley faces started being used in discussion forums in the 1980s to indicate that a comment is a joke [12], and people placing them to indicate laughter, thus making the discussion more closely resemble the patterns in oral, embodied communication [27].

When the basic character strings consisting of ASCII characters started giving away to the colorful pictorial images of emojis, created in Japan in the late 1990s [21], the expressive power of emojis increased. Emojis are used as a short way to convey pragmatic information through imagery [5], for aiding personal expression and reducing the ambiguity of discourse of the message [22], and as non-verbal communication cues [41]. The phenomenon of emoji has raised discussion about whether it is evolving to a visual language of its own, and its resemblance to early historical visual communication systems, such as Egyptian hieroglyphs, has been pointed out [5]. It has been argued that graphic emoticons will lead in the future to the emergence of a universal symbolic language [6],

especially considering face-based emojis, due to the universality of facial expressions [5]. While there has been much discussion on cultural differences in emoji use [24], this primarily applies to 'symbol' icons while there is relative similarity for 'smileys' and 'people' icons [14].

2.3 Emojis as User Feedback Tools

Aiming to tap into emojis' ability for emotional communication, emojis have been increasingly integrated into research tools exploring users' perceptions of products and services. The most prevalent approach has been the development of various *smiley face Likert scales*, typically constructed with icons of 5 faces with expressions ranging from sad to happy, e.g. [15, 29], see Figure 2. Alismail and Zhang [4] studied the use of smiley face Likert scales in online surveys, reporting no significant differences between numeric and emoji versions, but suggested using emoji-based questionnaires for assessing subjective feelings. Similarly, Toepoel et al. [36] found no difference in results from radio button and smiley face based Likert scales but reported that participants preferred the smiley face version. In earlier work, Alismail and Zhang [3] developed a version of the UEQ using smiley face Likert scales instead of the usual radio buttons. However, this implementation retained words to define the rating criteria and used the same happy-to-sad set of emojis for each of the subscales [3], see Figure 2.

Another approach to using emojis as a feedback mechanism is requiring participants to select one or more emojis from a pre-defined list of emojis, e.g., with use cases in the domains of orthodontics [26] and food preferences [19, 20]. Typically, these methods present a list of between 30 and 50 emojis from which the survey participants *click all that apply* (CATA) [26, 32], see Figure 2. In particular, a large body of work using this approach has focused on its use in assessing food preferences, e.g., [19, 20, 32]. When used in a mobile web survey with millenials as participants, Bosch and Revilla [8] reported a clear preference for answering open-ended questions with CATA-style emoji responses.

One step in developing emoji-based UX feedback tools is selecting emojis universally understood to represent the emotions or characteristics being evaluated. To determine the set of emojis to use in an online feedback tool Sun et al. [34] conducted an online survey where participants were presented with 2 or 3 expert-selected emojis and asked to rate how well each emoji represented a specific emotion. Similarly, Scherr et al. [30] curated a set of emojis for use in gathering feedback and noted that most participants had a very homogeneous understanding of the sentiments and emotions represented by the emojis.

2.4 Our Contribution

There has been a huge amount of prior research in the HCI domain on using emojis in conversational usage, particularly exploring differences in understanding the meaning of emojis, e.g., Togans et al. [37]. However, there has been relatively little research on emoji use as part of UX evaluation tools in HCI. In the HCI domain, prior work has almost exclusively focused on using smiley face Likert scales [4, 15, 29, 36]. As the smiley face Likert scale only uses five emojis, it does not leverage the capability of the larger set of emojis

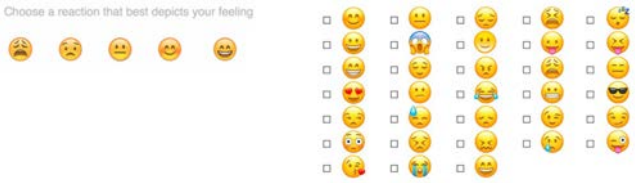

Figure 2: Prior work using emoji-based feedback. Left: A typical example of the ubiquitous happy-to-sad rating scale, the Emojiscore scale [1] is also used in [3, 4]. Right: Emoji-based click all that apply (CATA) questionnaires [32]. Alternative presentations of this approach include an emoji keyboard [8].

typically used in conversational settings to explore emotional perceptions of products and services. The use of a larger set of emojis for feedback has been little researched in HCI. Existing UX tools such as AttrakDiff and the UEQ have been systematically designed to provide assessment across the multiple aspects contributing to the overall UX, and are well-established tools in the HCI domain. It now seems timely to explore the bringing together of these threads, combining the word-based UEQ with the potential for emotional expression supported by an extended set of emojis. Our research introduces several novel approaches:

(1) Employing a diverse range of emojis in feedback scales, moving beyond the commonly used 5-emoji happy-to-sad scale.
(2) Utilizing emojis to represent the endpoints of bipolar rating scales.
(3) Adapting the UEQ, a systematically designed word-based UX feedback tool, to use pairs of emojis instead of adjective pairs.

3 MAPPING UEQ TERMS TO EMOJIS

The first step in developing an emoji-based UEQ was mapping individual UEQ terms to emojis that are understood to have a similar meaning. This was done by first creating a proposed set of word-emoji mappings that were then validated through a user study.

3.1 Defining Representative Emojis

To identify broadly understood emoji representations for the terms in the UEQ, a set of potential mappings was created, which would be validated in a user study. The primary criterion for matching a term with an emoji was the definition of the emoji provided by Emojipedia [13]. The emoji meanings on Emojipedia are written by emoji experts and lexicographers and include both the originally intended emoji meaning and the current real-world use [13]. Two researchers independently identified one or more emojis whose definition they considered best represented each term. A discussion phase followed, aiming to reach a consensus on the single best emoji representation. For example, for the UEQ term *boring* the yawning face emoji was selected as its definition includes "... to imply boredom with a person or topic" [13]. Similarly, *interesting* was mapped to face with monocle, defined as "... pondering, considering, or questioning something, ... or encouraging a closer look at some content" [13]. For terms such as *leading edge* this required a broader interpretive approach, e.g. Rocket emoji "... to

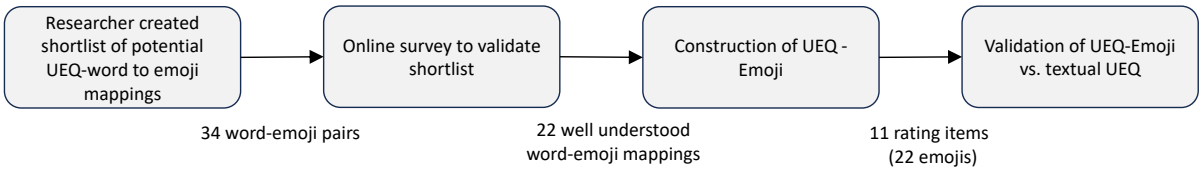

Figure 3: The iterative steps in the research process

the moon" [13] metaphorically representing technological advancement. As well as manually searching the meaning descriptions, the researchers used OpenAI's ChatGPT-3.5 to suggest potential mappings, which were then reviewed. For some terms, no potential emoji representations were identified.

With the exception of 2 emojis (*soon* and *prohibited*), all the selected emojis were ranked in the top 100 used emojis in 2021 [9]. Of the 26 rating items in the full UEQ, 15 were selected as having potential emoji representations. As each rating item consisted of 2 opposing terms this corresponded to 30 emojis. For the terms *usual* and *impractical* 2 alternative emojis were included, and for *interesting* 3 alternative emojis were included. Thus the proposal list contained 34 word-emoji pairs. The selected list aimed to provide full coverage of the 6 subscales in the UEQ (*Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty*), including at least 2 proposals in each subscale (see Table 2).

This list was then evaluated in an online survey to identify how well the proposed mappings were understood. This process was carried out in 2 phases, 1) using only the 16 terms in the 8-item short version of the UEQ (UEQ-S) and 2) including pairs of terms selected from the full 26-item version of the UEQ (total 52 terms).

3.2 User Evaluation of Word-Emoji Mappings

The user evaluation was conducted as an online survey, using participants recruited via Prolific [2]. The survey was divided into two phases, the first evaluating 16 word-emoji mappings, and the second a different set of 19 word-emoji mappings. 40 participants completed each survey. The participants in each phase were different. For each word-emoji pair, participants rated how well the emoji represented the word using a 7-point rating scale. The presentation order of the items was randomized to reduce bias effects.

Participants were selected as native English speakers, located in the UK who identified themselves as regular users of chat and messaging apps. This sample aimed to provide a participant set with a common understanding of the UEQ terms. Prior works translating the UEQ between different languages, e.g. German to Spanish, have noted that a one-to-one translation of the terms is not possible [28]. Participants' mean age was 37.6 years (sd = 13.5) with 74% identifying as female and 26% as male. The survey took a median time of 100 sec to complete, for which participants were compensated 0.50€. The compensation level was suggested by the recruitment platform and corresponded to an hourly rate of approximately 17€.

Results from the survey are presented in Figure 4.

4 CONSTRUCTING AND EVALUATING THE EMOJI-BASED UEQ

Based on the word-emoji mapping ratings, we then formed a set of bipolar rating scales.

4.1 Constructing the Emoji-based UEQ

The ratings for the individual word-emoji mappings were combined as opposite pairs, and the mean rating for each pair was calculated, see Table 2. The best-performing emoji was selected where several different emojis had been proposed for a word. Based on the pairs with mean ratings on the positive side, an emoji-based UEQ was created, including 11 rating items, see Figure 1.

Although the target was to develop a version of the UEQ that addressed all the subscales represented in the UEQ, providing a balanced view of user experience, this was not possible. The 11-item emoji UEQ does not include any items addressing the *novelty* subscale. The developed scale includes five items contributing to pragmatic quality and two contributing to hedonic quality [33]. Additionally, the four attractiveness items contribute to both pragmatic and hedonic quality.

4.2 User Study

An online survey was arranged to evaluate the emoji-based UEQ. We compared the developed 11-item emoji-UEQ and the same 11 items presented in word format in a within-subjects study design. The within-subjects design enabled participants to select their preferred design in an end questionnaire. Following the method used to develop the short UEQ (UEQ-S) [33], the question "What is your opinion of the company Amazon?" was used. This subject is likely well-known to all participants and enabled our findings to be compared with those reported for the UEQ-S design. The presentation order of word and emoji versions was randomized to reduce bias effects. As well as demographics, to identify participants' familiarity with emoji use, participants provided information on the number of emojis they typically use.

4.3 Survey Participants

Participants for the online survey were recruited using the Prolific platform [2]. Participants were selected as native English speakers located in the UK who identified themselves as regular users of chat and messaging apps. The survey took a median time of 2 m 5 s to complete, for which participants were compensated €0.80. The recruitment platform suggested the compensation level corresponding to an hourly rate of approximately €21. We excluded participants who provided the same rating to all scales in the questionnaire. We collected valid responses from 47 participants, 34

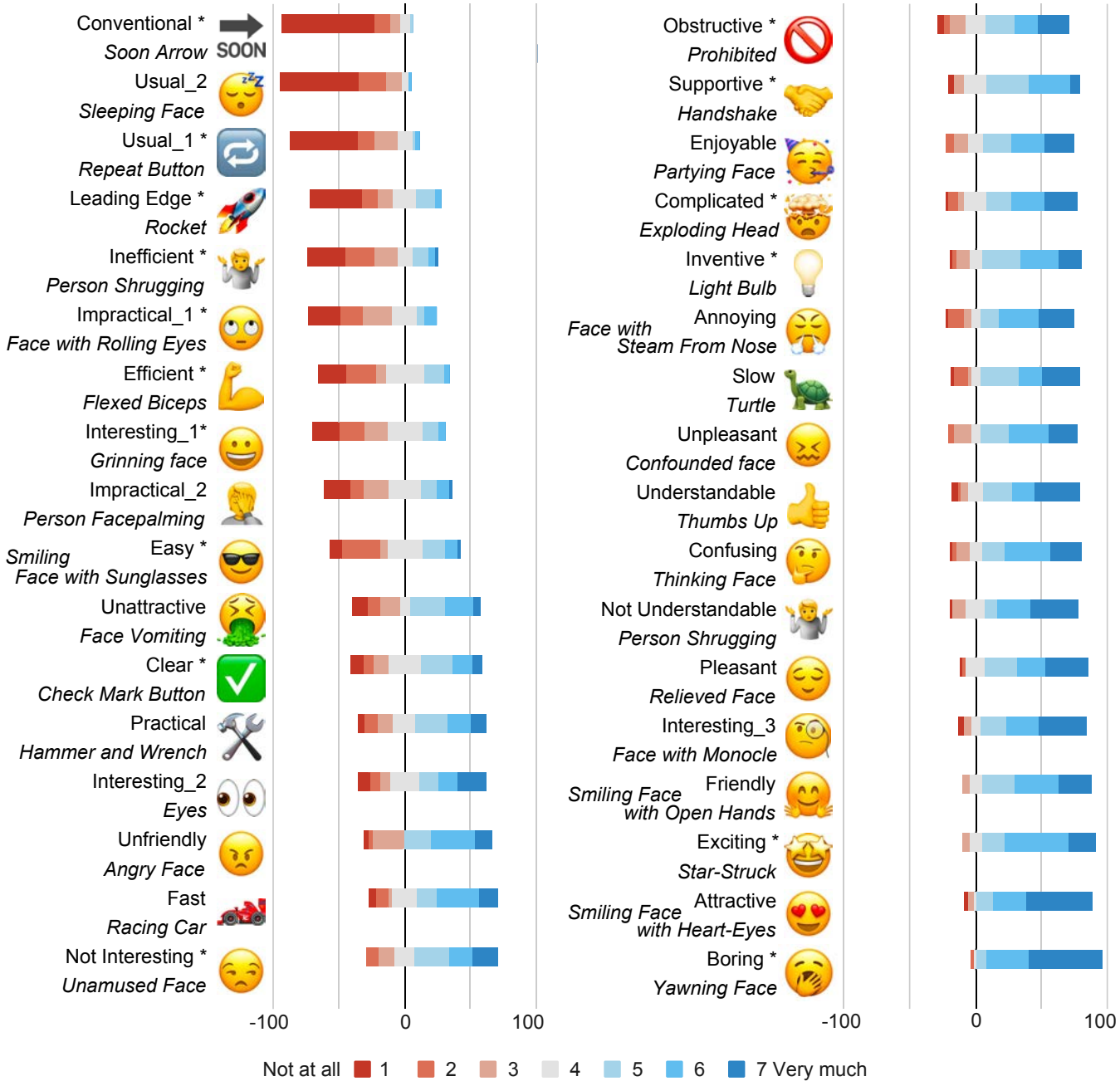

Figure 4: Ratings for how well the pre-selected emojis represent the UEQ terms. Items marked with an asterisk from study 1.

identifying as women and 13 as men. The participants' ages ranged from 19 to 77 years, with a mean age of 39.04 (SD = 13.74) and a median age of 37. Participants' range of emoji usage varied, with 20 participants reporting typically using between 0 and 5 different emojis, 19 using 6 - 10, and 8 between 11 - 30. Most participants (approximately 89%) completed the study on a mobile device.

4.4 Results

We present the survey's overall results in Figure 5. We first conducted a Shapiro-Wilk test for all measures and found that they were not normally distributed. Thus, we performed non-parametric tests in the following. Here, we report Wilcoxon signed-rank tests and additionally report the Bayesian version of the Wilcoxon signed-rank to highlight the likelihood that the emoji-UEQ results yield

Table 1: The 11 selected emoji pairs mapped to the six subscales in the UEQ.

Attractiveness	Perspicuity	Efficiency	Dependability	Stimulation	Novelty
Pleasant	Understandable	Fast	Supportive	Exciting	
Attractive	Easy			Interesting	
Friendly	Clear				
Enjoyable					

Table 2: Mean ratings and standard deviations for emoji pairs (1-7 scale) and related UEQ subscale

Emoji pair	Mean rating	SD	UEQ Subscale
Exciting / Boring	6.0	1.0	Stimulation
Understandable / Not Understandable	5.3	1.6	Perspicuity
Pleasant / Unpleasant	5.3	1.5	Attractiveness
Interesting / Not interesting	5.2	1.6	Stimulation
Friendly / Unfriendly	5.1	1.4	Attractiveness
Enjoyable / Annoying	5.0	1.7	Attractiveness
Attractive / Unattractive	5.0	1.6	Attractiveness
Supportive / Obstructive	4.9	1.6	Dependability
Fast / Slow	4.9	1.7	Efficiency
Clear / Confusing	4.8	1.6	Perspicuity
Easy / Complicated	4.3	1.6	Perspicuity
Practical / Impractical	3.9	1.7	Efficiency
Inventive / Conventional	3.4	1.3	Novelty
Efficient / Inefficient	2.9	1.6	Efficiency
Leading edge / Usual	2.4	1.5	Novelty

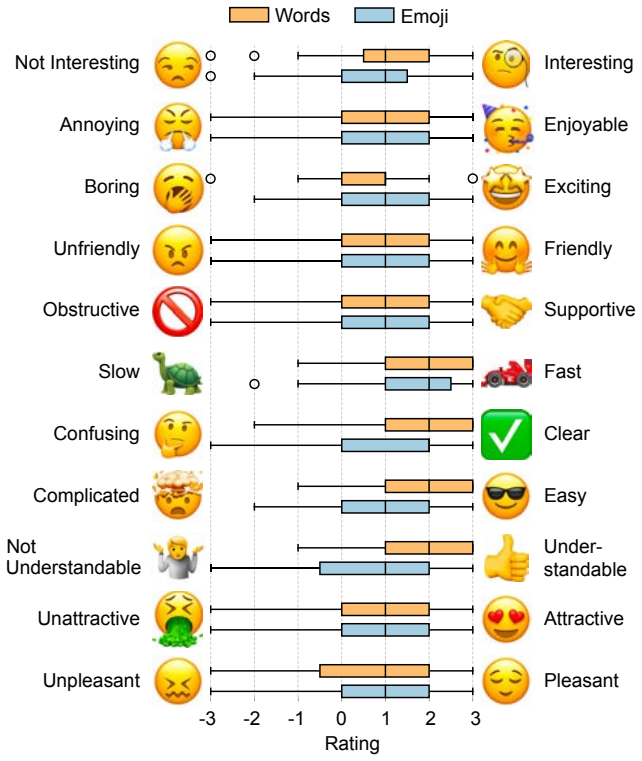

Figure 5: Final questionnaire responses "Opinion of the company Amazon" using word and emoji-based rating scales. Note the 1-7 rating scale has been transposed to -3 to +3 as described in Laugwitz et al. [23].

similar results to the original version, see Table 3. The analysis revealed significant differences between the emoji and word rankings for the scales *understandable*, *easy clear*, and *interesting*. In all cases, the ratings from the word-based condition were higher than those of the emoji condition. For the remaining 7 scales, no significant difference between the conditions was noted.

After checking for normality, we performed a Wilcoxon signed-rank test to compare the task completion time (TCT) answering word and emoji versions. The test revealed no significant difference in the times for words ($M = 30.76s$, $SD = 13.44s$) and emojis ($M = 30.24s$, $SD = 10.71s$); see Table 3. To complete each UEQ, 11 clicks (or touchscreen taps) were required, with the number of clicks increasing above this if participants adjusted their ratings before submitting the form. Again, we performed the Wilcoxon signed-rank test to compare the number of clicks for words and emojis. The results revealed no significant difference in the number of clicks for words ($M = 12.19$, $SD = 1.57$) and emojis ($M = 12.47$, $SD = 1.93$); see Table 3. To identify the effects of participants' current emoji usage, independent samples t-tests were conducted between participants who reported typically using 0-5 different emojis and participants who reported using a wider selection of emojis. The tests revealed no significant differences in the scores for words ($t = -0.68$, $p = .50$) or emojis ($t = 0.50$, $p = .62$). This suggests that higher familiarity with emojis did not significantly affect the rating of the formats.

Based on participants' ratings for ease of use of the two UEQ formats, a paired t-test was conducted. The test revealed a significant difference in the scores for words ($M = 6.68$, $SD = 0.69$) and emojis ($M = 5.57$, $SD = 1.47$); $t = 4.54$, $p < .001$, Cohen's d = 0.96. Hence, participants considered the word-based format significantly easier to use than the emoji format, and this effect was large.

Table 3: Comparison of word-based and emoji-based ratings using Wilcoxon signed-rank and Bayesian alternative methods (See Figure 5). * We tested the normality using the Shapiro-Wilk test.

	Wilcoxon signed-rank test			Bayesian Alternative		Normality*	
	W	z	p	W	BF10	W	p
UEQ Overall	213.5	−2.976	.003	213.5	23.334	.947	<.001
Pleasant	143.5	.552	.574	143.5	.208	.875	<.001
Attractive	156.5	.186	.857	156.5	.165	.909	<.001
Understandable	84.	−3.366	<.001	84.	64.734	.870	<.001
Easy	88.	−3.718	<.001	88.	505.427	.898	<.001
Clear	93.	−2.505	.011	93.	4.044	.954	<.001
Fast	35.5	−1.941	.05	35.5	.584	.850	<.001
Supportive	242.5	.899	.355	242.5	.257	.911	<.001
Friendly	209.5	.493	.613	209.5	.201	.919	<.001
Exciting	264.5	1.814	.059	264.5	.749	.935	<.001
Enjoyable	137.5	−.357	.712	137.5	.185	.897	<.001
Interesting	71.	−2.835	.004	71.	31.483	.920	<.001
TCT	616.	.55	.589	616.	.167	.819	<.001
Clicks	357.	1.367	.161	357.	.392	.718	<.001

5 DISCUSSION

In this section, we discuss further validation of our developed tool, highlight what we consider to be its strengths and weaknesses and reflect on the limitations of our design and research process.

5.1 Validating the UEQ-Emoji

When creating new evaluation scales, it is common to assess the internal validity of items that contribute to the same scale, e.g., using Cronbach's alpha. For the text-based UEQs, this internal validity check was a key part of the tool's validation process, e.g., [33]. In the case of the UEQ-Emoji, we present Cronbach's alpha values for the Attractive, Practical, and Hedonic scales in Table 4. According to standard interpretations of Cronbach's alpha, these values indicate excellent internal consistency. These measures of internal validity are consistent with those reported for the text-based UEQ-S, which had Cronbach's alpha values of 0.85 for pragmatic quality and 0.81 for hedonic quality [33].

In the original validation of the UEQ-S, the authors compared the mean values of its scales to those of the full UEQ for the same question, rating the company Amazon. They reported a mean value of 1.17 for Practical quality (which includes Efficiency, Perspicuity, and Dependability scales) and a mean value of 0.66 for Hedonic quality (comprising Stimulation and Originality scales) [33]. As our study utilized the same question, rating the company Amazon, it is also interesting to compare with our findings – noting that opinions may have changed in the six years between the studies. Our findings showed similar values: 1.04 for Practical quality and 0.73 for Hedonic quality. In the words of the UEQ-S creators, the UEQ-Emoji "seems to approximate" to the word-based UEQs "expectedly well" [33].

For further validation, the creators of the UEQ conducted a user study that revealed a negative correlation between the time taken to complete a test task and participants' ratings on the UEQ's perspicuity scale. Interestingly, this correlation was not observed on the hedonic side [23, p. 9]. As future work, a similar study could be completed with the UEQ-Emoji.

Table 4: Internal consistency of scales in the UEQ-Emoji measured by Cronbach's Alpha

Category	Scales	Cronbach's Alpha
Attractive	Pleasant, Attractive, Enjoyable, Friendly	0.962
Practical Quality	Understandable, Easy, Clear, Fast, Supportive	0.885
Hedonic Quality	Exciting, Interesting	0.897

5.2 Strengths and Weaknesses of the UEQ-Emoji

The value of emojis in communicating emotions when used in communication is well-reported [39]. While the scope of our study does not explicitly deliver evidence on the benefits of emojis over word-based UX evaluation tools, this has been consistently reported by numerous prior works. For example, exploring survey responses with freely selected emojis Bosch and Revilla [8] reported that the average information conveyed was significantly higher than words. We believe these benefits will also be reflected in our use of emojis.

One possible advantage of the UEQ-Emoji is its suitability for mobile use. Traditional word-based rating scales, like AttrakDiff and the UEQ, often struggle to fit on a smartphone screen in portrait mode, especially when bipolar scales with labels at both ends are used. This issue is further complicated by the lengthy labels often found in such tools. For example, the UEQ-S has been translated into over 36 languages, and in most Latin-based languages, the labels are too long to fit without requiring horizontal scrolling to read them fully, see Figure 6. In contrast, the UEQ-Emoji can be displayed in usable form on a smartphone screen in portrait orientation, enhancing its usability for mobile users. To emphasize the importance of mobile optimization, we note that 90% of the participants in our online survey used a smartphone rather than a laptop or desktop computer. However, despite the non-optimal visual presentation, participants using smartphones still expressed a significant preference for the text-based format. In a future study, it would be interesting to collect data identifying if participants scrolled the screen sideways to read the labels or guessed the positive label without it being fully visible.

We note that our study participants preferred the word-based tool over the emoji version when rating ease of use. We consider that the relatively old age of our sample, with a median age of 37 years, may have influenced this stated preference. As a comparison, Schouteten and Meiselman [31] evaluated children's food preferences using an emoji-based feedback tool and reported on the tool's potential with this user segment.

A significant limitation of our UEQ-Emoji is its inability to capture the perceived novelty of the product or service being evaluated,

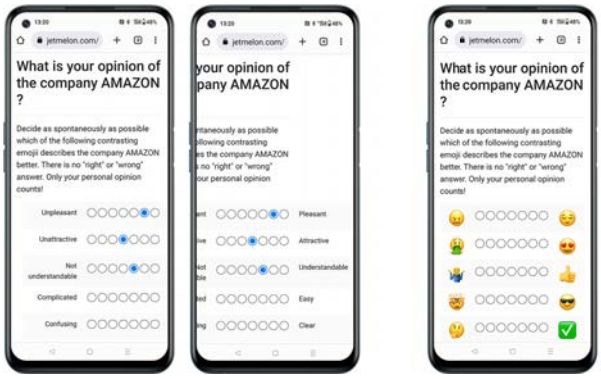

Figure 6: The existing text-based UEQ and the UEQ-Emoji displayed on a smartphone. With a readable font size and usable size rating selection buttons, the text-based version does not fit in portrait but requires horizontal scrolling

a feature in the text-based UEQ. This is due to the semantic limitations of the emoji language, as no emojis are available that are universally understood to mean *conventional* or *leading edge*. We suspect these terms in the word-based UEQ, are also subject to varied interpretations. It would be interesting to explore the performance of this scale in the word-based UEQ.

Several other challenges have been previously noted with emoji-based questionnaires; a useful list is provided by Alismail and Zhang [3]. For instance, some participants in prior studies felt that an emoji-based questionnaire appeared unprofessional. Others speculated that the emojis might introduce a positive bias in the results, a phenomenon we also observed to some extent in our study. The need for additional cognitive effort when interpreting emojis, akin to translating a second language, was also highlighted, leading to longer completion times for emoji-based questionnaires [8]. In our study, we did not observe significant differences in the time taken to complete the survey. We note that the paid online participants in our study are likely motivated to complete the survey in the least amount of time possible.

Given these mixed strengths and weaknesses, further research is essential to understand the potential and limitations of the UEQ-Emoji fully. In particular, it would be interesting to utilize the tool with different demographic groups, such as children, who may have a different interaction experience.

5.3 Limitations of our Work

Our developed UEQ-Emoji is based on the well-established UEQ. However, reflecting on the impact of our process of selecting the used emojis for our final UEQ-Emoji is valuable. In particular, its initial stage required the researchers to pre-select a set of potential emoji-word mappings. We believe that the possibility of researcher bias in this step is minimal. The process required matching the meaning of UEQ terms with the defined meanings of emojis, i.e., semantic similarity. Focusing primarily on the top 100 most used emojis, 34 were selected for validation in a user study. We note that similar processes were used in selecting emojis for emoji-based choose all that apply (CATA) tools [19, 20, 32].

6 CONCLUSION

This paper has presented the design and initial validation study of the UEQ-Emoji, an emoji-based tool for evaluating user experience. The developed tool evaluates both pragmatic and hedonic aspects of the overall user experience. Following the text-based UEQ tools, the UEQ-Emoji includes scales measuring *attractiveness*, *perspicuity*, *efficiency*, *dependability*, and *stimulation*. However, due to the semantic limitations of the emoji language, no measure for *novelty* is included.

The meaning of the emojis used in the UEQ-Emoji was validated to ensure they were understood to have the same meaning as the terms in the text-based UEQ. Based on a limited evaluation (n = 47), the final version of the UEQ-Emoji shows satisfactory internal consistency and a reasonable degree of comparability to the established text-based UEQ. In a within-subjects study with participants with a median age of 37, the traditional word-based evaluation tool was preferred over the UEQ-Emoji. Further research is needed to fully understand the potential benefits of the UEQ-Emoji in the HCI context. Specifically, we speculate that the tool may be particularly beneficial when used with younger populations, such as children and adolescents. We provide a ready-to-use UEQ-Emoji form and analysis tool as supplementary materials to this paper.

ACKNOWLEDGMENTS

This research has been supported by the Academy of Finland as part of the AWARE project (Grant number: 355694) and the Lapland Robotics project, funded by the European Regional Development Fund (ERDF).

REFERENCES

[1] 2023. Emojiscore. https://emojiscore.com/. Accessed: 12th October 2023.
[2] 2023. Prolific · Quickly find research participants you can trust. https://www.prolific.co/. Accessed: 12th October 2023.
[3] Sarah Alismail and Hengwei Zhang. 2018. The use of emoji in electronic user experience questionnaire: An exploratory case study. In *Proceedings of the 51st Hawaii International Conference on System Sciences*. University of Hawai'i, Hawaii, USA, 10. http://hdl.handle.net/10125/50315
[4] Sarah Alismail and Hengwei Zhang. 2020. Exploring and Understanding Participants' Perceptions of Facial Emoji Likert Scales in Online Surveys: A Qualitative Study. *Trans. Soc. Comput.* 3, 2, Article 12 (apr 2020), 12 pages. https://doi.org/10.1145/3382505
[5] Hamza Alshenqeeti. 2016. Are emojis creating a new or old visual language for new generations? A socio-semiotic study. (2016), 14 pages. https://ssrn.com/abstract=3709343
[6] Junichi Azuma. 2012. Graphic emoticons as a future universal symbolic language. In *Audiovisual translation and media accessibility at the crossroads*. Brill, Leiden, The Netherlands, 61–84. https://doi.org/10.1163/9789401207812_006
[7] Qiyu Bai, Qi Dan, Zhe Mu, and Maokun Yang. 2019. A systematic review of emoji: Current research and future perspectives. *Frontiers in psychology* 10 (2019), 2221. https://doi.org/10.3389/fpsyg.2019.02221
[8] Oriol J Bosch and Melanie Revilla. 2021. Using emojis in mobile web surveys for Millennials? A study in Spain and Mexico. *Quality & Quantity* 55, 1 (2021), 39–61. https://doi.org/10.1007/s11135-020-00994-8
[9] Unicode Consortium. 2023. Emoji Frequency. https://home.unicode.org/emoji/emoji-frequency/ Accessed: 12th October 2023.
[10] Anna DeCastellarnau. 2018. A classification of response scale characteristics that affect data quality: a literature review. *Quality & quantity* 52, 4 (2018), 1523–1559. https://doi.org/10.1007/s11135-017-0533-4
[11] Ignacio Díaz-Oreiro, Gustavo López, Luis Quesada, and Luis A. Guerrero. 2019. Standardized Questionnaires for User Experience Evaluation: A Systematic Literature Review. *Proceedings* 31, 1 (2019), 12. https://doi.org/10.3390/proceedings2019031014
[12] Eli Dresner and Susan C. Herring. 2010. Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Communication theory* 20, 3 (2010), 249–268. https://doi.org/10.1111/j.1468-2885.2010.01362.x

[13] Emojipedia. 2023. Emojipedia — Home of Emoji Meanings. https://emojipedia.org/ Accessed: 12th October 2023.

[14] Sharath Chandra Guntuku, Mingyang Li, Louis Tay, and Lyle H Ungar. 2019. Studying cultural differences in emoji usage across the east and the west. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 226–235.

[15] Lynne Hall, Colette Hume, and Sarah Tazzyman. 2016. Five Degrees of Happiness: Effective Smiley Face Likert Scales for Evaluating with Children. In *Proceedings of the The 15th International Conference on Interaction Design and Children* (Manchester, United Kingdom) *(IDC '16)*. Association for Computing Machinery, New York, NY, USA, 311–321. https://doi.org/10.1145/2930674.2930719

[16] HappyOrNot. 2023. Stop guessing. Get customer feedback platform that works | HappyOrNot®. https://www.happy-or-not.com/ Accessed: 12th October 2023.

[17] Marc Hassenzahl and Noam Tractinsky. 2006. User experience-a research agenda. *Behaviour & information technology* 25, 2 (2006), 91–97. https://doi.org/10.1080/01449290500330331

[18] Jan Karem Höhne, Dagmar Krebs, and Steffen-M Kühnel. 2022. Measuring income (in) equality: Comparing survey questions with unipolar and bipolar scales in a probability-based online panel. *Social Science Computer Review* 40, 1 (2022), 108–123. https://doi.org/10.1177/0894439320902461

[19] Sara R Jaeger, Christina M Roigard, David Jin, Leticia Vidal, and Gastón Ares. 2019. Valence, arousal and sentiment meanings of 33 facial emoji: Insights for the use of emoji in consumer research. *Food research international* 119 (2019), 895–907. https://doi.org/10.1016/j.foodres.2018.10.074

[20] Sara R Jaeger, Yixun Xia, Pui-Yee Lee, Denise C Hunter, Michelle K Beresford, and Gastón Ares. 2018. Emoji questionnaires can be used with a range of population segments: Findings relating to age, gender and frequency of emoji/emoticon use. *Food Quality and Preference* 68 (2018), 397–410. https://doi.org/10.1016/j.foodqual.2017.12.011

[21] Linda K Kaye, Stephanie A Malone, and Helen J Wall. 2017. Emojis: Insights, affordances, and possibilities for psychological science. *Trends in cognitive sciences* 21, 2 (2017), 66–68. https://doi.org/10.1016/j.tics.2016.10.007

[22] Linda K Kaye, Helen J Wall, and Stephanie A Malone. 2016. "Turn that frown upside-down": A contextual account of emoticon usage on different virtual platforms. *Computers in Human Behavior* 60 (2016), 463–467. https://doi.org/10.1016/j.chb.2016.02.088

[23] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*. Springer Berlin Heidelberg, Berlin, Heidelberg, 63–76. https://doi.org/10.1007/978-3-540-89350-9_6

[24] Xuan Lu, Wei Ai, Xuanzhe Liu, Qian Li, Ning Wang, Gang Huang, and Qiaozhu Mei. 2016. Learning from the Ubiquitous Language: An Empirical Analysis of Emoji Usage of Smartphone Users. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 770–780. https://doi.org/10.1145/2971648.2971724

[25] Simon Makin. 2023. People Differ Widely in Their Understanding of Even a Simple Concept Such as the Word 'Penguin'. *Scientific American* (2023). https://www.scientificamerican.com/article/people-differ-widely-in-their-understanding-of-even-a-simple-concept-such-as-the-word-penguin1/ Accessed: 12th October 2023.

[26] Guido Artemio Marañón-Vásquez, Lucianne Cople Maia, Luisa Schubach da Costa Barreto, Mariana Farias da Cruz, Lucas Alves Jural, Mônica Tirre de Souza Araújo, and Matheus Melo Pithon. 2022. Emoji as promising tools for emotional evaluation in orthodontics. *Progress in Orthodontics* 23, 1 (2022), 28. https://doi.org/10.1186/s40510-022-00418-3

[27] Robert R. Provine, Robert J. Spencer, and Darcy L. Mandell. 2007. Emotional Expression Online: Emoticons Punctuate Website Text Messages. *Journal of language and social psychology* 26, 3 (2007), 299–307. https://doi.org/10.1177/0261927X06303481

[28] Maria Rauschenberger, Siegfried Olschner, Manuel Perez Cota, Martin Schrepp, and Jörg Thomaschewski. 2012. Measurement of user experience: A Spanish language version of the user experience questionnaire (UEQ). In *7th Iberian Conference on Information Systems and Technologies (CISTI 2012)*. IEEE, 1–6.

[29] Janet C Read, Stuart MacFarlane, and Chris Casey. 2002. Endurability, Engagement and Expectations: Measuring Children's Fun. (2002), 23 pages.

[30] Simon Scherr, Svenja Polst, Lisa Müller, Konstantin Holl, and Frank Elberzhager. 2019. The Perception of Emojis for Analyzing App Feedback. *International Journal of Interactive Mobile Technologies (iJIM)* 13, 13 (2019), 18. https://doi.org/10.3991/ijim.v13i02.8492

[31] Joachim J Schouteten and Herbert L Meiselman. 2021. The potential of using emoji instead of words when measuring emotions associated with food. *Food Quality and Preference* 92 (2021), 104182. https://doi.org/10.1016/j.foodqual.2021.104182

[32] Joachim J Schouteten, Jan Verwaeren, Sofie Lagast, Xavier Gellynck, and Hans De Steur. 2018. Emoji as a tool for measuring children's emotions when tasting food. *Food quality and preference* 68 (2018), 322–331. https://doi.org/10.1016/j.foodqual.2018.03.005

[33] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 6 (2017), 103–106. https://doi.org/10.9781/ijimai.2017.09.001

[34] Na Sun, Elise Lavoué, Chulakorn Aritajati, Aurélien Tabard, and Mary Beth Rosson. 2019. Using and perceiving emoji in design peer feedback. In *13th International Conference on Computer Supported Collaborative Learning (CSCL'19)*. International Society of the Learning Sciences (ISLS), Lyon, France, 8. https://repository.isls.org//handle/1/4417

[35] Hamed Taherdoost. 2019. What is the best response scale for survey and questionnaire design; review of different lengths of rating scale/attitude scale/Likert scale. (2019), 10 pages. https://ssrn.com/abstract=3588604

[36] Vera Toepoel, Brenda Vermeeren, and Baran Metin. 2019. Smileys, stars, hearts, buttons, tiles or grids: influence of response format on substantive response, questionnaire experience and response time. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 142, 1 (2019), 57–74. https://doi.org/10.1177/0759106319834665

[37] LaCount J Togans, Thomas Holtgraves, Gyeongnam Kwon, and Tania E Morales Zelaya. 2021. Digitally saving face: An experimental investigation of cross-cultural differences in the use of emoticons and emoji. *Journal of pragmatics* 186 (2021), 277–288. https://doi.org/10.1016/j.pragma.2021.09.016

[38] Arnold P. O. S. Vermeeren, Effie Lai-Chong Law, Virpi Roto, Marianna Obrist, Jettie Hoonhout, and Kaisa Väänänen-Vainio-Mattila. 2010. User Experience Evaluation Methods: Current State and Development Needs. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries* (Reykjavik, Iceland) *(NordiCHI '10)*. Association for Computing Machinery, New York, NY, USA, 521–530. https://doi.org/10.1145/1868914.1868973

[39] Sarah Theres Völkel, Ramona Schoedel, Lale Kaya, and Sven Mayer. 2022. User Perceptions of Extraversion in Chatbots after Repeated Use. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 253, 18 pages. https://doi.org/10.1145/3491102.3502058

[40] Joseph B Walther and Kyle P D'addario. 2001. The impacts of emoticons on message interpretation in computer-mediated communication. *Social science computer review* 19, 3 (2001), 324–347. https://doi.org/10.1177/089443930101900307

[41] Yiyi Yang. 2020. Are you emoji savvy? Exploring nonverbal communication through emojis. *Communication Teacher* 34, 1 (2020), 2–7. https://doi.org/10.1080/17404622.2019.1593472