Usability and Emotions of Mental Health Assessment Tools: Comparing Mobile App and Paper-and-Pencil Modalities

Yang S. Liu[ab], Jeffrey Hankey[c], Nigel Mantou Lou[d], Pratap Chokka[c], and Jason M. Harley[bde]*

[a]*Department of Psychiatry, University of Alberta, Edmonton, Canada*

[b]*Department of Educational Psychology, University of Alberta, Edmonton, Canada*

[c]*Chokka Center for Integrative Health, Edmonton, Alberta, Canada*

[d]*Research Institute of the McGill University Health Centre, Montreal, Canada*

[e]*Department of Surgery, McGill University Health Centre, Montreal, Canada*

*corresponding author email: jason.harley@mcgill.ca

Funding and acknowledgement

We acknowledge funding from the MITACS Accelerate program (Grant Number: IT10022 and IT10918) as well as the Chokka Center for Integrative Health for additional funding support, research site provision and onsite support.

Usability and Emotions of Mental Health Assessment Tools: Comparing Mobile App and Paper-and-Pencil Modalities

Users' experiences in mental health assessment are multifaceted, including their emotional experiences. Yet, studies of mobile apps for psychiatric assessment have centered on diagnostic accuracy and perceived usability, with little consideration of the impact of user emotional experiences. In this study, we focused on users' perceived usability and emotions and compared the user experience of a paper-and-pencil and an app-based collection of mental health screening questionnaires: EarlyDetect. The System Usability Scale (SUS) and modality-directed emotion questionnaires were administered using paper-and-pencil or iPad. Modality was assigned pseudo-randomly on patients' first visit at a referral-based mental health clinic. We found that patients assigned to the iPad app reported a significantly higher SUS score than patients assigned to paper-and-pencil, qualified by a modality-by-gender interaction where modality effects were significant for men but not for women. Moreover, enjoyment was positively linked to perceived usability, whereas boredom, frustration, and anxiety were negatively linked to usability. Our findings illustrate the added value of studying user experience applied to psychiatric assessments, where both emotions and gender-specific user experience should be taken into consideration. We further discuss the implications for psychiatric assessments via app versus traditional data collection.

Keywords: usability; mobile app; emotion; m-health; mental health assessment

Introduction

Usability of Mental Health Assessment: Why It Matters.

About 25% of people worldwide are influenced by mental health issues and/or neurological disorders at some point in their lives (World Health Organization, 2001). In Canada alone, where this study was conducted, Mental health issues affect seven million Canadians at the cost of over $50 billion per year (Mental Health Commission of Canada, 2015). In Canada, the 12-month and life-time prevalence rates of Major Depressive Disorder (MDD) alone are 4.7% and 11.2%, respectively (Knoll &

MacLennan, 2017). While prevention of depression is possible with early intervention (Almeida, 2014; Hall & Reynolds-lili, 2014), general practitioners fail to correctly identify depression in up to half of the cases (Mitchell, Vaze & Rao, 2009).

One solution is to enable screening assessments using mobile apps, a computer program or software application designed to run on a mobile handheld computing device such as a smartphone or tablet. With 2.8 billion unique smartphone subscribers worldwide (GSMA Intelligence, 2017), and 27 million (72%) Canadians using smartphones (Newzoo, 2018), mobile technologies represent a promising platform for e-health applications. Accordingly, mental health app development is exploding, receiving wide acceptance among users and clinicians (Areàn, Hoa, & Andersson, 2016; Bradford & Rickwood, 2015). Technology-facilitated mental health interventions and services are widely available (Andersson, 2016; Barak & Grohol, 2011; Epstein et al., 2017; Hollis et al., 2018) and their clinical efficacy has been demonstrated on mobile app-based products (Bakker, Kazantzis, Rickwood, & Rickard, 2018; Berger, Krieger, Sude, Meyer, & Maercker, 2017). Diagnosis of psychiatric conditions relies, however, on the physician (Bilello, 2016), creating a bottleneck for patients waiting to receive treatment. This study contributes to the knowledge base surrounding the user experience of digital screening tools, as technology-facilitated mental health screening is a promising, yet under-developed and under-studied, area to expedite initial mental health diagnosis and treatment (Maunder & Hunter, 2018; Ospina-Pinillos, Davenport, Ricci, Milton, Scott, & Hickie, 2018; van Bebber, Meijer, Wigman, Sytema, & Wunderink, 2018).

From a practical point of view, mobile app based mental health assessment is more efficient and has many advantages over other modalities. (Bakker et al., 2018; Berger et al., 2017; Bradford & Rickwood, 2014). Advantages include, but are not

limited to: (1) saving time for patients in filling out forms by reducing the number of questions to be answered by using response-adaptive, decision tree navigation, (2) providing faster reporting, so that doctors can get immediate assessment results without manual entry, and (3) providing more accurate information, such as the use of automatic timestamps, reminders of missing responses, and storing information directly to reduce manual scoring and processing errors.

However, Hassenzahl (2018) noticed that some practitioners might worry about the incompatibility between usability (e.g., time saving) and emotional experiences (e.g., enjoyment), as research showed that usability is associated with emotional responses. For example, users experienced more positive emotions when they perceived a webpage to be more concise (vs. complex; Goldberg, 2012). As such, a practical product may not necessarily produce positive emotion. In other words, usability and emotion are two distinct constructs (cf. Harley et al., 2019), which may have different implications for understanding users' experiences. Therefore, it is important to examine both users' perceived usability and emotional experiences. Moreover, from a users' point of view, it is unclear whether the user would experience an additional psychological burden using an app for mental health assessment. For example, it is not known whether negative emotions such as frustration, anxiety, and boredom could be reduced or exacerbated by the testing modality (i.e., app vs. paper-and-pencil), and whether a low level of negative emotions would be found across testing modalities. The assessment of emotional experiences in a mental health app is particularly important. Patients who go to a clinic may have certain levels of negative emotions because they are not healthy and are seeking mental health assessment and treatment. That is, users of mental health-related applications are usually people who experience more negative emotions than others. Therefore, It is crucial to ensure that the assessment modality

itself does not exacerbate already existing negative emotions.

Perceived Usability and Emotions of Mental Health Assessment: Mobile App Vs. Paper-And-Pencil

In this study, we compared users' perceived usability and emotional experiences of a mobile app mental health assessment with a traditional paper-and-pencil mental health assessment. The use of mobile technology (e.g., mobile applications) is growing in different types of assessments, including screening patients' mental health. Before mobile apps were popular, internet-browser-based assessments had already been introduced and studied. Studies comparing an internet-browser-based modality with a paper-and-pencil modality reported mixed results. Although some found modality equivalency regarding the accuracy of assessment (e.g., Cronly et al. 2018; Riva, Teruzzi & Anolli, 2004), other studies reported between modality differences. Specifically, compared to paper-and-pencil, patients who filled out an assessment using an internet-browser-based modality scored higher on the Beck Depression Inventory (Schulenberg & Yutrzenka, 1999). Although similar in concept to the Internet-browser-based modality, studies with mobile apps have been focused on scores and psychometric properties (e.g., Buchanan et al., 2005; Buchanan, Johnson, & Goldberg, 2005, Cronly et al. 2018; Schulenberg & Yutrzenka, 1999) and user behaviours (Weigold, Weigold, Drakeford, Dykema, & Smith, 2016; Weigold, Weigold & Natera, 2018) rather than the user experience.

We argue that the mobile app interface is inherently different from the traditional computer-based modality due to its touch input, different operating systems, and smaller screen size. Usability, defined as the extent to which a product can be utilized by users to achieve specified goals with effectiveness, efficiency, and satisfaction in a particular context (Bevan, Carter, Harker, 2015; ISO, 1998), is a core

construct of user experience, focusing on the reliable measurement of quality of interaction with the product. Mobile apps have demonstrated high perceived usability according to research using the System Usability Scale (SUS; Brooke, 1996), where Kortum and Sorber (2015) reported the average SUS score of 15 high-usage mobile apps across Android and iOS platforms to be 77.7, equivalent to a "B+" on Sauro & Lewis (2016)'s Curved Grading Scale (CGS), putting usability at a comparable level to other consumer products such as Wii, DVRs, GPS, and PowerPoint. For comparison, a score of 68 is a "C" on CGS and considered to be "acceptable" (Bangor, Kortum, & Miller, 2009).

In addition to the quality of interaction, user experience includes users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviours, and accomplishments that occur before, during, and after use (International Organization for Standardization [ISO], 2010; Coursaris, & Kim, 2011; Partala & Kangaskorte, 2009). User experience cannot be simply reported via usability and it is also essential to study emotions (Agarwal & Meyer, 2009). The emotional design model (Norman, 2003) suggests that users' cognition (e.g., what they think about the app) and emotions (e.g., how they feel about using the app) are processed through different psychological mechanisms and should therefore both receive attention in user experience research. Emotion is an essential aspect of user experience (Jeon, 2017), which can arise in response to a variety of stimuli, including appraisals of the goal-conduciveness of technology (Harley, Poitras, Jarrell, Duffy, & Lajoie, 2016, Harley, Lajoie, Tressel, & Jarrell, 2020; Harley, Liu et al., 2019; Poitras, Harley, & Liu, 2019). Moreover, Hassenzahl (2018) argued that practitioners should embrace the concept of users' subjective emotional experiences in their designs.

However, usability measurements are often inconsistent in practice, including the measurement of emotions (Borsci, Federici, Malizia, & De Filippis, 2019). For example, studies on user experience have mainly focused on positive emotions (Hassenzahl & Tractinsky, 2006). A systematic review of user experience showed that only 13 out of 180 studies included negative emotions in their measures (Hornbæk, 2006). Understanding negative emotions and what predicts negative emotions can provide implications to help further improve user experience (Partala & Kangaskorte, 2009; Partala & Kallinen, 2012). Emotional states have also been shown to impact cognitive processes (Jarrell, Harley, Lajoie, & Naismith, 2017; Pekrun, 2006; Pekrun & Perry, 2014), where positive emotions such as enjoyment foster motivation and promote situationally-appropriate information processing, whereas negative emotions such as boredom undermine motivation and interest. In addition, privacy concerns may influence emotions, as the use of the internet has been associated with fear of privacy invasion, leading to computer-induced negative emotions (Schulenberg & Yutrzenka, 1999). This concern might be exacerbated when patients are inputting sensitive information on a device that is not their own. However, it is unclear whether negative emotions such as frustration, anxiety, and boredom could be reduced or exacerbated by the testing modality, and whether a low level of negative emotions would be found across testing modalities.

The Current Study

In this study, we explored the user experience of an app by focusing on the usability and emotional impact of EarlyDetect, a mental health assessment tool that contains multiple, clinically validated questionnaires (Authors, 2021). EarlyDetect is available in both mobile app and paper-and-pencil formats. The primary aim of this study was to explore whether the mobile app modality facilitated a better user

experience, as measured by higher usability and higher positive and lower negative emotions than the paper-and-pencil modality. To address our study goal, we asked two research questions:

(RQ1) Does perceived usability significantly differ between testing modalities? We hypothesized that the mobile app testing modality would evoke higher SUS scores than paper and pencil, as suggested by previous findings (Fuller-Tyszkiewicz, 2018). Second, we hypothesized that gender would not have an effect on usability across modality, based on results from 206 usability tests (Bangor, Kortum & Miller, 2008).

(RQ2) Do self-reported emotions significantly differ between testing modalities? We hypothesized that participants using the mobile app testing modality would report no significant differences in emotional experience compared to those using paper-and-pencil. That is, the app would not induce an additional emotional burden to users. Specifically, we expected that both groups of participants would experience low levels of enjoyment and anger/frustration based on the nature of the tasks (i.e., filling out a questionnaire). We also expected medium levels of anxiety and boredom because participants were filling out a questionnaire about their mental health which may arouse some anxiety and is unlikely to be enjoyable. Finally, based on research regarding computer-induced negative emotions (Schulenberg & Yutrzenka, 1999), we expected that the app might be slightly, but not significantly higher for anxiety on account of it being a less familiar modality.

(RQ3) Is self-reported usability significantly correlated with emotional experiences? Previous research suggests that perceived usability is associated with positive emotional responses (Goldberg, 2012). Therefore, we hypothesized that perceived usability would be positively correlated with positive emotional responses and negatively correlated with negative emotional responses.

EarlyDetect Questionnaires

EarlyDetect was developed to assist physicians to better diagnose mental health illnesses by (1) leveraging integrated, multiple clinical questionnaires in a standardized format, and (2) offering a broader range of mental health assessment approximating a clinical interview process (Authors, 2021). The questionnaires includes (in testing order): (a) Life History Questionnaire (LHQ; proprietary—see supplementary materials), (b) Mini International Neuropsychiatric Interview (M.I.N.I)-Major Depressive Episode (Lecrubier et al., 1997), (c) Patient Health Questionnaire (PHQ-9; Kroenke & Spitzer, 2002), (d) M.I.N.I-(Hypo) Manic Episode, (e) The Mood Disorder Questionnaire (MDQ; Miller, Klugman, Berv, Rosenquist, & Ghaemi, 2004), (f) M.I.N.I-Generalized Anxiety Disorder, (g) Generalized Anxiety Disorder (GAD-7; Spitzer, Kroenke, Williams, & Löwe, 2006), (h) ADHD Screening (proprietary), (i) Adult ADHD Self-Report Scale (ASRS-v1.1 – Part A) Symptom Check List (Kessler et al., 2005), (j) M.I.N.I – Alcohol Abuse and Dependence, (k) Alcohol Use Disorders Identification Test (AUDIT; Babor, Higgins-Biddle, Saunders, Monteiro, 2001), and (l) Sheehan Disability Scale (SDS; Sheehan, Harnett-Sheehan & Raj, 1996).

EarlyDetect is response-adaptive and it streamlines how questions are asked and how questions are presented on the mobile interface to minimize the burden for participants to answer unnecessary questions. Specifically, the app applies decision rules to skip asking more detailed questions while assessing a specific mental health condition if patients screened negative on the *M.I.N.I.* screening questions. For example, for bipolar disorder screening, the patient was asked “Have you ever had a period of time when you were feeling ‘up’, ‘high’ or ‘hyper’ or so full of energy or full of yourself that you got into trouble, or that other people thought you were not your usual self? (Do not consider times when you were intoxicated on drugs or alcohol)”. If a patient indicated they do not have any of the symptoms of mania in a stemming

question, they wound not need to answer the follow-up questions from MDQ that probe for more details. Figure S1a in supplementary materials provides another example, where patient's answers on the *M.I.N.I. - Major Depressive Episode* will determine if the app presents the PHQ-9 to patient subsequently.

Methods

Participants and Procedure

Over a period of four months, 191 participants provided informed consent upon arrival at their first visit at an interdisciplinary, referral-based mental health clinic serving an ethnoculturally diverse population in a large urban centre in western Canada. Participation was voluntary, with no monetary incentives, and approved by the University of Alberta ethics board. Participants were given an information package and consent form by clinic staff. They completed the EarlyDetect questionnaires followed by the user experience questionnaires (see supplementary materials). Participants completed the questionnaires using either paper-and-pencil or the mobile app. The process took approximately 10 to 15 minutes, after which participants underwent a clinical interview with a psychiatrist for assessment and diagnosis. Both assessment modalities contained the same sets of questionnaires presented in the same sequential navigational structure. However, the presentation was not qualitatively identical due to the inherent affordances and constraints between modalities (see Figure S1 in supplementary materials for an illustration). For example, compared to writing responses on a piece of paper, interacting with an iPad involves tapping and swiping fingers, such that the presentation of contents was optimized according to the modality. The modality assignment was pseudo-random. Participants used the paper-and-pencil version if (a) they were randomly assigned to this modality; b) if all iPads were

occupied; or c) they were offered an iPad but requested to use paper-and-pencil instead. In total, seven of the 191 participants were excluded due to incomplete data (see supplementary materials for details). Overall, 73 participants used the app version (35 female, M_{age} = 32.0) and 118 used the paper-and-pencil version (71 female, M_{age} = 36.3).

Questionnaires

The System Usability Scale (SUS)

The SUS has been applied to previous mobile app usability studies (Korturm & Sober, 2015), demonstrating a high overall score of 77.7, and Cronbach α of 0.88, across ten apps. Fuller-Tyszkiewicz (2018) also applied SUS to an e-mental health app, reporting a high usability score of 86. Our app-specific version of SUS has replaced the keyword "system" with the "EarlyDetect questionnaire" from the original version (Brooke, 1996). Cronbach's α indicated that the scale reliability was high for SUS (α = .83)

Emotion

To evaluate modality-directed emotion, questionnaires from Harley and colleagues (Harley et al., 2016, 2018, 2019) and Poitras and colleagues (2019) were modified to include experiences of enjoyment, boredom, frustration, and anxiety. Each emotion was measured using a single item to avoid item fatigue (Harley et al., 2015). The questionnaire used a five-point Likert scale where 1 corresponded to strongly disagree and 5 to strongly agree. Participants reported their modality-directed emotions, which required them to think about their feelings directed toward the testing modality. For example: "I enjoyed using the paper-based version of these surveys". Emotions

were correlated in the expected directions, providing a measure of internal validity (Pekrun & Perry, 2014; see Table 1).

Data analysis

For SUS, we conducted a two-way analysis of covariance (ANCOVA) to compare the effect of testing modality (paper, app) and gender (male, female) while controlling for age because prior research suggested age is associated with the acceptance of smartphone assessment (Ramsey et al., 2016). For emotion analyses, we conducted Multivariate analysis of covariance (MANCOVA) on all emotions to help protect against inflating the Type-1 error rate in the follow-up ANCOVAs and post-hoc comparisons. The emotions were moderately correlated, suggesting the MANCOVA's assumption of no multicollinearity were met (Table 1). The MANCOVA tested for mean differences and interactions between modality and gender for modality-directed emotions, while controlling for age and SUS. A series of two-way ANCOVAs on each of the four dependent variables was conducted following the MANCOVA, with Bonferroni adjusted post-hoc comparisons.

- Insert Table 1 here-

Results

(RQ1) Does usability significantly differ between testing modalities?

In line with our hypothesis, usability, as measured by SUS, was higher for the app than paper modality (M_{app} = 75.87, M_{paper} = 69.89, $t(182)=2.43$, $p < 0.05$, d =0.37). As shown in Figure 1, a two-way ANCOVA analysis revealed a significant main effect of modality on SUS scores, $F(1,179) = 4.70$, $p < .05$, partial eta-squared (η_{p2}) = .03, qualified by a significant interaction effect between modality and gender, $F(1,179) = 6.19$, $p < .05$, $\eta_p^2 = .03$) (Table 1). Post-hoc analysis showed that male patients had a

significantly ($p < .05$) higher SUS score for app than paper testing modality. In contrast, female patients showed no significant SUS score difference between modalities.

- Insert Figure 1 here-

(RQ2) Do self-reported emotions significantly differ between testing modalities?

A statistically significant MANCOVA test was obtained for SUS, Pillai's Trace = .18, $F(4,175) = 9.68$, $p < .05$, $\eta_p^2 = .18$, Modality, Pillai's Trace = .05, $F(4,175) = 2.47$, $p < .05$, $\eta_p^2 = .05$, and a modality-by-gender interaction, Pillai's Trace = .15, $F(4,175) = 7.47$, $p < .05$, $\eta_p^2 = .15$. A main effect of SUS score was statistically significant for all emotions with effect sizes (η_p^2) ranging from .06 (enjoyment) to .11 (anxiety). A main effect of modality was statistically significant for anxiety $F(1,178) = 4.02$, $p < .05$, $\eta_p^2 = .02$. Modality-by-gender interactions were statistically significant for enjoyment and boredom, with effect sizes (η_p^2) ranging from .03 (enjoyment) to .08 (boredom) (Table 1).

A series of post-hoc analyses for the ANCOVA's between-modality differences (mobile vs. paper-and-pencil) found that men reported lower levels of boredom and higher levels of anxiety in the mobile app condition compared to paper-and pencil. In contrast, women reported lower levels of enjoyment and higher levels of boredom and frustration, for the app modality compared to paper-and-pencil. Despite the gender and modality differences, all negative modality-directed emotions were reported at a low level, i.e., $M < 3.00/5.00$.

(RQ3) Is self-reported usability correlated with emotional experiences?

We found that perceived usability was correlated with different emotional experience at a moderate level (Table 1). Specifically, perceived usability is significantly positively correlated with enjoyment ($r = .24$, $p < .01$) and significantly

negatively correlated with boredom (r = -.35, p < .01), frustration (r = -.36, p < .01), and anxiety (r = -.31, p < .01).

Discussions

In this study, we found the mobile app testing modality yielded an overall higher SUS score than the paper-and-pencil testing modality. Although our results are limited to iPad, the usability of phone apps is consistently higher than iPad apps when the app is identical (Kortum & Sorber, 2015). Our result, therefore, offers preliminary support for the potential generalizability of the modality's main effect. However, contrary to our hypothesis, this effect is qualified by a gender-by-modality interaction, such that only male patients, but not female patients, showed significant SUS score differences between modalities. These findings diverged from Kortum and Sorber's (2015) results and suggested that the mobile app assessment may benefit males more than females regarding the improvement of their overall user experiences. However, future studies need to include both iPad and phone modality to facilitate interpretation of the gender-by-modality interaction.

Consistent with computer-induced negative emotions (Schulenberg & Yutrzenka, 1999), the app modality had a small effect (η_{p2} = .03) on inducing modality-directed anxiety. However, we found modality-directed negative emotions are gender-specific, where women show more boredom and frustration and less enjoyment toward the app modality compared to paper-and-pencil. In contrast, men showed less boredom but higher anxiety toward the app modality compared to paper-and-pencil. These results highlight that emotions are associated with user experience, but that emotions arise from more than just appraisals related to user experience, consistent with ISO's (2010) definition of user experience. For example, participants might have been reflecting on how filling out mental health information on a particular modality made them feel.

Despite the gender and modality differences, all negative modality-directed emotions were reported at a low level. Given the demonstrated clinical efficacy of digital mental health interventions, we see this as preliminary evidence that mobile app implementation of these specific tools could be a user-friendly and environmentally-friendly alternative to paper-and-pencil.

We contend that the higher usability of the mobile app modality complements other practical benefits of mobile app assessments, including saving patients' time, providing faster reporting, and reducing errors. Further, the results of our study provide preliminary insight into how emotions interact with usability. In particular, negative emotions such as boredom and frustration were all negatively associated with usability, whereas enjoyment was positively associated with usability. This association pattern between usability and emotions was previously found in Harley et al.'s (2019) study using the Edmonton Queer History mobile app, where both SUS and Emotion questionnaires were administered, suggesting negative emotions can potentially be reduced if usability is enhanced. Future research should explore this direction.

Limitations and Future Research

Limitations of this study include pseudo-random group assignment: in some cases, (a) all iPads were occupied or (b) participants requested paper-and-pencil instead of the iPad (see supplementary materials), thus preventing us from making strong claims with respect to potential causal relationships between study variables. However, self-selection (i.e., preference for pencil and paper) was rare (estimated at 2-3 from clinical staff), and we expect the effect size in favour of iPad may have been slightly larger if self-selection had not been possible. Another limitation is that we only used iPads, limiting our claims on whether usability differences can be attributed to EarlyDetect, not iPads. Additionally, we had a medium rather than a large sample size,

and effect sizes were small. To overcome these limitations, future randomized experiments may include a larger sample size that randomly assign participants to different types of modalities and assess technology acceptance or comfort levels to understand the causal effects of mobile apps on users' experiences. Moreover, although this study revealed that male participants (but not female participants) perceived higher usability of the app (vs. paper-based), we did not know why female participants did not benefit from the mobile app regarding their usability. Future qualitative studies that draw on semi-structured interviews could play an important role in better understanding such gender differences (cf. Vermeeren et al., 2010). Relatedly, future research using a mixed-methods study approach (cf. O'Brien & Lebow, 2013) can also provide a more comprehensive view of the usability and emotional experiences of EarlyDetect.

Regarding measurements, we used self-report measurements of emotion and usability, which are subject to biases. Nevertheless, the self-report emotional questionnaire used in this study has been validated in previous studies and found to be have strong internal consistency and external validity: correlations within emotions were in the expected directions and the correlations between emotions and performance outcomes in other settings were also in an expected direction (e.g., Harley et al., 2016, 2019, 2020; Poitras et al., 2019). There are many factors that may affect users' experiences (Koursaris & Kim, 2011). In addition to gender that we focused on in this study, future research may include the type of mental health issue the patients reported, educational levels, and technology fluency to understand how these factors may interact with patients' experience. Finally, due to this study being conducted in a clinic with patients rather than with a convenience sample of undergraduate students, we were limited in the scope and follow-up of questions we could reasonably ask. With a different population, future studies should examine the baseline of emotion to fully

understand whether and how pre-existing emotions may interact with users' emotions in using the app or paper-and-pencil assessment. However, researchers and practitioners should be mindful of the challenges to pinpoint emotional experiences because of their complex dynamic nature, which can change momentarily due to different personal and situational factors (Harley, Pekrun, Taxer, & Gross, 2019).

Practical Implications

Our findings showed that user experience in the EarlyDetect mobile app is better than the traditional data collection method in general. We recommend the use of the app modality for user screening/data collection and future studies in similar settings. However, the use of mobile apps in data collection is not without challenges from a user experience perspective. Researchers and practitioners should be cognizant of factors such as user gender, testing modality, and the potential interactions between factors (e.g., a gender by testing modality interaction; cf. Harley et al., in press). In our study, the positive effect of the mobile app (vs. paper-based) method on user experience was found only among male participants. The mobile app did not improve but neither did it undermine females' experiences. Future research should study how to improve females' user experiences. Finally, users' experiences are multifaceted and involve different emotional responses (Partala & Kangaskorte, 2009). Researchers and practitioners who wish to understand users' emotional responses can use the four-item emotion scale to understand users' enjoyment, boredom, frustration, and anxiety (see appendix). This scale (Harley et al., 2016, 2019, 2020; Poitras et al., 2019) is associated with users' perceived usability. Specifically, enjoyment was positively linked to perceived usability, whereas boredom, frustration, and anxiety were negatively linked to usability.

Conclusion

This paper presented a study investigating the perceived usability and emotional experience with incoming patients at a referral-based psychiatric clinic. First, our results supported the use of mobile apps, such as EarlyDetect, in mental health screening from a user experience perspective. We found that the app is associated with a higher SUS score than the paper-and-pencil testing modality. However, this benefit was found only among male patients, but not female patients. For female participants, the mobile app did not improve or undermine their experience compared to those in the paper-and-pencil condition. Therefore, we recommend that future usability studies should further examine whether gender differences can be found in other apps or modalities. This is important because overgeneralization to all users may exacerbate biases in the data collection process. Future studies may examine factors or use methods that may help identify areas of dissatisfaction in females' user experience (e.g., using semi-structured interviews) in order to improve females' perceptions of usability.

Second, we highlight that users' emotional experiences are linked to their perceived usability, such that people who reported higher usability levels also experienced more positive emotions. As such, our finding that females showed higher levels of boredom and frustration in the app vs. paper-and-pencil condition finding resonates with previous studies that females reported higher levels of negative technology-related emotion (e.g., Cai, Fan, & Du, 2017). As such, it is not surprising that compared to males, females showed little improvement in their perceived usability in the app (vs. the paper-and-pen). Emotional experiences provide an additional layer to understand user experiences and gender differences. Therefore, we recommend the inclusion of emotional experiences to comprehensively understanding user experience in future studies (cf. Berkman & Karahoca, 2016).

References:

Authors. (2021).

Agarwal, A., & Meyer, A. (2009). Beyond usability: evaluating emotional response as an integral part of the user experience. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems* (pp. 2919-2930).

Almeida, O. P. (2014). Prevention of depression in older age. *Maturitas*, *79*(2), 136-141.

Andersson, G. (2016). Internet-delivered psychological treatments. Annual Review of Clinical Psychology, 12, 157-179.

Areàn, P., Hoa, L.K., & Andersson, G. (2016). Mobile technology for mental health assessment. *Dialogues Clinical Neuroscience*, *18*(2), 163-169.

Babor T.F., Higgins-Biddle, J. C., Saunders, J. B., Monteiro, M. G. AUDIT: the alcohol use disorders identification test: guidelines for use in primary health care. Geneva: World Health Organization; 2001.

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2018). A randomized controlled trial of three smartphone apps for enhancing public mental health. *Behaviour research and therapy*, *109*, 75-83.

Bangor, A., Kortum, P., & Miller, J. A. (2008). The system usability scale (SUS): An empirical evaluation. *International Journal of Human-Computer Interaction*, *24*(6), 574-594.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, *4*(3), 114-123.

Barak, A., & Grohol, J. M. (2011). Current and future trends in internet-supported mental health interventions. *Journal of Technology in Human Services*, *29*(3), 155-196.

Barak, A., Hen, L., Boniel-Nissim, M., & Shapira, N. (2008). A Comprehensive Review and a Meta-Analysis of the Effectiveness of Internet-Based Psychotherapeutic Interventions. *Journal of Technology in Human Services*, *26*(2-4), 109-160.

Barua B. Fraser Institute. 2015. Waiting your turn: wait times for health care in Canada. Retrieved from: https://www.fraserinstitute.org/sites/default/files/waiting-your-turn-2015.pdf

Berger, T., Krieger, T., Sude, K., Meyer, B., & Maercker, A. (2017). Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: Results of a pragmatic randomized controlled trial. *Journal of Affective Disorders*, *227*, 455-462.

Berkman, M. I., & Karahoca, D. (2016). Re-assessing the usability metric for user experience (UMUX) scale. *Journal of Usability Studies*, *11*(3), 89-109.

Bevan, N., Carter, J., & Harker, S. (2015, August). ISO 9241-11 revised: What have we learnt about usability since 1998? In *International Conference on Human-Computer Interaction* (pp. 143-151). Springer, Cham.

Bilello, J. A. (2016). Seeking an objective diagnosis of depression. *Biomarkers in Medicine*, *10*(8), 861-875.

Borsci, S., Federici, S., Malizia, A., & De Filippis, M. L. (2019). Shaking the usability tree: why usability is not a dead end, and a constructive way forward. *Behaviour & Information Technology*, 38(5), 519-532.

Bradford, S., & Rickwood, D. (2014). Electronic psychosocial assessment tool: Concept development and identification of barriers to successful implementation. *Journal of Technology in Human Services*, *32*(4), 275-296.

Bradford, S., & Rickwood, D. (2015). Acceptability and utility of an electronic psychosocial assessment (myAssessment) to increase self-disclosure in youth mental healthcare: a quasi-experimental study. *BMC Psychiatry*, *15*(1), 305.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4-7.

Buchanan, T., Ali, T., Heffernan, T. M., Ling, J., Parrott, A. C., Rodgers, J., & Scholey, A. B. (2005). Nonequivalence of on-line and paper-and-pencil psychological tests: The case of the prospective memory questionnaire. *Behavior Research Methods*, *37*(1), 148-154.

Buchanan, T., Johnson, J. A., & Goldberg, L. R. (2005). Implementing a five-factor personality inventory for use on the internet. *European Journal of Psychological Assessment, 21*, 115–127.

Cai, Z., Fan, X., & Du, J. (2017). Gender and attitudes toward technology use: A meta-analysis. *Computers & Education*, *105*, 1-13.

Cronly, J., Duff, A. J., Riekert, K. A., Perry, I. J., Fitzgerald, A. P., Horgan, A., ... & Savage, E. (2018). Online versus paper-based screening for depression and anxiety in adults with cystic fibrosis in Ireland: a cross-sectional exploratory study. *BMJ Open, 8*(1)

Coursaris, C. K., & Kim, D. J. (2011). A meta-analytical review of empirical mobile usability studies. Journal of usability studies, 6(3), 117-171.

Goldberg, J. H. (2012). Relating perceived web page complexity to emotional valence and eye movement metrics. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 501-505). Sage CA: Los Angeles, CA: SAGE Publications.

Epstein, R., Ho, M., Hyun, S., Le, C., Robertson, R. E., & Stout, D. (2017). A DSM-5-Based Online Mental Health Referral Inventory: A Large-Scale Validation Study. *Journal of Technology in Human Services*, *35*(3), 231-246.

Fuller-Tyszkiewicz, M., Richardson, B., Klein, B., Skouteris, H., Christensen, H., Austin, D., ... & Shatte, A. (2018). A Mobile App–Based Intervention for Depression: End-User and Expert Usability Testing Study. *JMIR Mental Health*, *5*(3), e54.

GSMA Intelligence. (2017). Number of unique mobile subscribers worldwide hits five billion. Retrieved from https://www.gsmaintelligence.com/

Hahs-Vaughn, D. L. (2016). *Applied multivariate statistical concepts*. Routledge.

Hassenzahl, M., Tractinsky, N. (2006). User experience – a research agenda. *Behavior & Information Technology, 25*, 91-97.

Hassenzahl, M. (2018). A personal journey through user experience. *Journal of Usability Studies, 13*(4), 168-176.

Hall, C. A., & Reynolds-Iii, C. F. (2014). Late-life depression in the primary care setting: challenges, collaborative care, and prevention. *Maturitas*, *79*(2), 147-152.

Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior*, *48*, 615-625.

Harley, J. M., Lajoie, S. P., Tressel, T., & Jarrell, A. (2020). Fostering positive emotions and history knowledge with location-based augmented reality and tour-guide prompts. *Learning and Instruction*, *70*, 101163

Harley, J. M., Liu, Y., Ahn, B. T., Lajoie, S. P., & Grace, A. P. (2020). Examining physiological and self-report indicators of empathy during learners' interaction

with a queer history app. *British Journal of Educational Technology, 51*(6), 1920-1937.

Harley, J. M., Liu, Y., Ahn, T. B., Lajoie, S. P., Grace, A. P., Haldane, C., ... & McLaughlin, B. (2019). I've got this: Fostering topic and technology-related emotional engagement and queer history knowledge with a mobile app. *Contemporary Educational Psychology, 59,* 101790.

Harley, J. M., Lou, N. M., Liu, Y., Cutumisu, M., Daniels, L. M., Leighton, J. P., & Nadon, L. (in press). University students' negative emotions in a computer-based examination: the roles of trait test-emotion, prior test-taking methods and gender. *Assessment & Evaluation in Higher Education*, 1-17.

Harley, J. M., Pekrun, R., Taxer, J. L., & Gross, J. J. (2019). Emotion regulation in achievement situations: An integrated model. *Educational Psychologist, 54*(2), 106-126.

Harley, J. M., Poitras, E. G., Jarrell, A., Duffy, M. C., & Lajoie, S. P. (2016). Comparing virtual and location-based augmented reality mobile learning: emotions and learning outcomes. *Educational Technology Research and Development, 64*(3), 359-388.

Hornbæk K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human–Computer Studies, 64*,79-102.

Hollis, C., Falconer, C. J., Martin, J. L., Whittington, C., Stockton, S., Glazebrook, C., & Davies, E. B. (2017). Annual Research Review: Digital health interventions for children and young people with mental health problems–a systematic and meta-review. *Journal of Child Psychology and Psychiatry*, *58*(4), 474-503.

International Organization for Standardization. (1998). *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability* (ISO/DIS 9241-11:1998). Retrieved from: https://www.iso.org/standard/16883.html

International Organization for Standardization. (2010). *Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems* (ISO/DIS 9241-210). Retrieved from: https://www.iso.org/standard/52075.html

Jeon, M. (2017). Emotions and affect in human factors and human–computer interaction: Taxonomy, theories, approaches, and methods. In *Emotions and affect in human factors and human-computer interaction* (pp. 3-26). Academic Press.

Jarrell, A., Harley, J. M., Lajoie, S., & Naismith, L. (2017). Success, failure and emotions: examining the relationship between performance feedback and emotions in diagnostic reasoning. *Educational Technology Research and Development*, *65*(5), 1263-1284.

Kessler, R.C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., Howes, M.J., Jin, R., Secnik, K., Spencer, T., Ustun, T.B., Walters, E.E. (2005). The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. *Psychological Medicine, 35*, 245-256.

Knoll, A. D., & MacLennan, R. N. (2017). Prevalence and correlates of depression in Canada: Findings from the Canadian Community Health Survey. *Canadian Psychology/ Psychologie Canadienne, 58*(2), 116-123.

Kroenke. K, & Spitzer, R.L. (2002) The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*, *32*, 1-7.

Kortum, P., & Sorber, M. (2015). Measuring the usability of mobile applications for phones and tablets. *International Journal of Human-Computer Interaction*, *31*(8), 518-529.

Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., ... & Dunbar, G. C. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *European psychiatry*, *12*(5), 224-231.

Maunder, R. G., & Hunter, J. J. (2018). An Internet Resource for Self-Assessment of Mental Health and Health Behavior: Development and Implementation of the Self-Assessment Kiosk. *JMIR Mental Health*, *5*(2), e39.

Mental Health Commission of Canada. (2015). *Informing the future: Mental health indicators for Canada.* Ottawa.

Mitchell, A. J., Vaze, A., & Rao, S. (2009). Clinical diagnosis of depression in primary care: a meta-analysis. *The Lancet*, *374*(9690), 609-619.

Miller, C. J., Klugman, J., Berv, D. A., Rosenquist, K. J., & Ghaemi, S. N. (2004). Sensitivity and specificity of the Mood Disorder Questionnaire for detecting bipolar disorder. *Journal of Affective Disorders*, *81*(2), 167-171.

Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist, 59*, 150–162.

Newzoo (2018). *Global mobile market report*. Retrieved from https://resources.newzoo.com

Norman, D. A. (2003). *Emotional design: Why we love (or hate) every things*. New York, NY: Basic Books.

O'Brien, H. L., & Lebow, M. (2013). Mixed-methods approach to measuring user experience in online news interactions. *Journal of the American Society for Information Science and Technology*, *64*(8), 1543-1556.

Ospina-Pinillos, L., Davenport, T. A., Ricci, C. S., Milton, A. C., Scott, E. M., & Hickie, I. B. (2018). Developing a mental health eClinic to improve access to and quality of mental health care for young people: Using participatory design as research methodologies. *Journal of Medical Internet Research*, *20*(5).

Partala, T., & Kangaskorte, R. (2009). The combined walkthrough: measuring behavioral, affective, and cognitive information in usability testing. *Journal of Usability Studies*, *5*(1), 21-33.

Partala, T., & Kallinen, A. (2012). Understanding the most satisfying and unsatisfying user experiences: Emotions, psychological needs, and context. *Interacting With Computers*, *24*(1), 25-34.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315-341.

Pekrun, R., & Linnenbrink-Garcia, L. (2014). Introduction to emotions in education. In *International handbook of emotions in education* (pp. 11-20). Routledge.

Poitras, E. G., Harley, J. M., & Liu, Y. S. (2019). Achievement emotions with location-based mobile augmented reality: An examination of discourse processes in simulated guided walking tours. *British Journal of Educational Technology*, *50*(6), 3345-3360.

Ramsey, A. T., Wetherell, J. L., Depp, C., Dixon, D., & Lenze, E. (2016). Feasibility and acceptability of smartphone assessment in older adults with cognitive and

emotional difficulties. *Journal of Technology in Human Services, 34*(2), 209-223.

Sauro, J., & Lewis, J. R. (2016). *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann.

Schulenberg, S. E., & Yutrzenka, B. A. (1999). The equivalence of computerized and paper-and-pencil psychological instruments: Implications for measures of negative affect. *Behavior Research Methods, Instruments & Computers, 31*, 315-321.

Schwarz, N., Strack, F., Hippler, H., & Bishop, G. (1991). The impact of administration mode on response effects in survey measurement. *Applied Cognitive Psychology, 5*(3), 193-212.

Sheehan, D.V., Harnett-Sheehan, K., Raj, B.A. (1996). The measurement of disability. *International Clinical Psychopharmacology, 11*(3), 89–95

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166,* 1092–1097.

van Bebber, J., Meijer, R. R., Wigman, J. T., Sytema, S., & Wunderink, L. (2018). A smart screening device for patients with mental health problems in primary health care: development and pilot study. *JMIR Mental Health, 5*(2), e41.

Vermeeren, A. P., Law, E. L. C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010, October). User experience evaluation methods: current state and development needs. In *Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries* (pp. 521-530).

Weigold, A., Weigold, I. K., Drakeford, N. M., Dykema, S. A., & Smith, C. A. (2016). Equivalence of paper-and-pencil and computerized self-report surveys in older adults. *Computers in Human Behavior, 54,* 407-413.

Weigold, A., Weigold, I. K., & Natera, S. N. (2018). Mean scores for self-report surveys completed using paper-and-pencil and computers: A meta-analytic test of equivalence. *Computers in Human Behavior, 86,* 153-164.

World Health Organization. (2001). *The World Health Report 2001: Mental health: new understanding, new hope.* World Health Organization.

Table 1

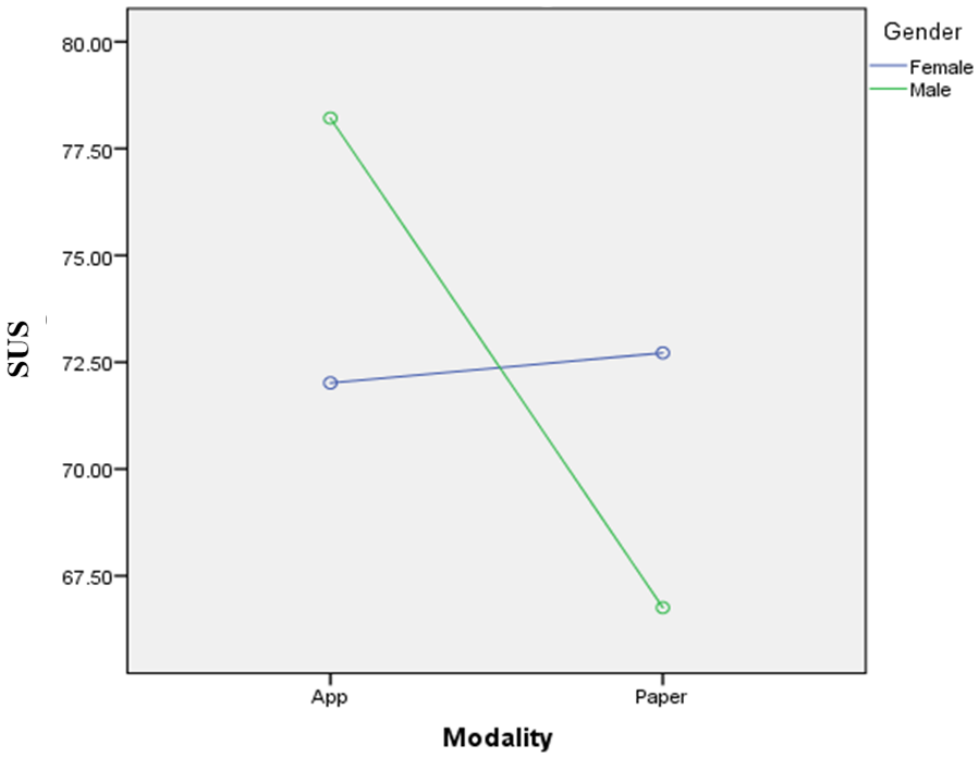
Correlations, descriptive statistics and analysis results of modality-directed emotions and usability

Pearson correlation and descriptive statistics							ANCOVAs						Post-hoc comparisons[a]					
							Modality		Gender		Modality×Gender		Female			Male		
Variable	*M*	*SD*	1	2	3	4	*F*	η_p^2	*F*	η_p^2	*F*	η_p^2	M_{app}	M_{paper}	*F*	M_{app}	M_{paper}	*F*
1.SUS	72.2	16.5	1				4.7*	.03	.00	.00	6.2*	.03	72.0	72.7	.04	78.2	67.8	10.4**
2.Enjoy	3.82	1.55	.24**	1			2.69	.02	.15	.00	6.28**	.03	3.30	4.25	9.19**	3.79	3.59	.33
3.Bored	2.38	1.43	-.35**	-.43**	1		.47	.00	3.92*	.02	16.2**	.08	2.58	1.92	5.89*	2.16	3.10	10.3**
4.Frustration	2.41	1.50	-.36**	-.41**	.59**	1	1.14	.01	.00	.00	15.1**	.08	3.00	1.96	13.1**	2.20	2.79	3.64
5.Anxiety	2.66	1.61	-.31**	-.13	.29**	.40**	4.02*	.02	2.90	.02	1.67	.01	2.61	2.43	.29	3.31	2.52	5.11*
N	184												35	67		37	45	

Note. *. significant at p < .05, **. significant at p < .01. *M* denotes mean, *SD* denotes standard deviation. N denotes number of sample used. Degrees of freedom for *F* statistics: F_{SUS} = (1, 179); $F_{emotions}$ = (1, 178); $F_{posthoc}$ = (1, 178).

Figure 1

Significant Interaction effect of gender and Modality on SUS



Biographical Notes

Dr. Yang S. Liu earned his PhD in Psychology at the University of Alberta, Canada, in 2015. Yang is a postdoctoral research follow at the department of psychiatry and the department of educational psychology at the University of Alberta.

Dr. Jeffrey Hankey earned his PhD in Educational Psychology at the University of Alberta, Canada, in 2020. Jeffrey is a research scientist at the Chokka Center for Integrative Health, Edmonton, Alberta, Canada.

Dr. Nigel Mantou Lou received his Ph.D. in Psychology at the University of Alberta in 2019. Nigel is a post-doc fellow at the Research Institute of the McGill University Health Centre.

Dr. Pratap Chokka is the founder and CEO of Chokka Center for Integrative Health, a clinical professor of psychiatry at the University of Alberta and a Consultant Psychiatrist at the Grey Nuns Hospital in Edmonton, Alberta, Canada.

Dr. Jason M. Harley is an assistant professor in the Department of Surgery, McGill University, a Junior Scientist at the Research Institute of the McGill University Health Centre, Director of the Simulation, Affect, Innovation, Learning, and Surgery (SAILS) Laboratory, and Associate Member of the Department of Educational and Counselling Psychology, McGill.

Supplementary materials

Additional details for methods

Two iPads with EarlyDetect installed were available in our clinic, and when both iPads were occupied by patients, the paper-and-pencil version was offered to participants. Because there's no blinding to the study, patients may have become aware the EarlyDetect was also available in a paper-and-pencil modality. A few patients self-selected the paper-and-pencil, but which patient self-selected paper-and-pencil were not formally documented. Our front staff estimated the frequency to be "very few people, about 1 in 30".

Of the 191 participants, seven of whom were excluded due to ambiguous or incomplete data. Five participants with 8 or more missing SUS answers were dropped from analyses. Fifteen participants with 3 or less missing SUS answers were retained with missing answers replaced by group mean, based on either the paper-and-pencil or app modality. Missing Emotion questionnaire answers from one participant were replaced by group mean. Two participants, one for each modality, who identified their gender as "other/prefer not to say" were removed due to the ambiguity of the gender, and statistical requirements of the gender analyses.

Of the 184 participants retained for analysis, 72 participants used the App version (Female = 35, M_{age} = 32.21) and 112 participants used the Paper version (Female = 67, M_{age} = 37.30).

For the multivariate analysis, Levene's test and residual normality checks were carried out and the ANCOVA assumptions were met.

EMOTION QUESTIONNAIRE FOR APP USERS

Emotions About App Usage

Interacting with apps can induce different feelings. This survey refers to the emotions you may have experienced when using the Early Detect App. Specifically, the following questions pertain to feelings you experienced WHILE using the app today. Record your answers below, using the appropriate number, where 1 indicates that you strongly disagree with a statement and 5 indicates that you strongly agree.

Questions:

I enjoyed using the EarlyDetect App.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Using the EarlyDetect App bored me.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Using the EarlyDetect App annoyed me.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Using the EarlyDetect App made me feel anxious.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

EMOTION QUESTIONNAIRE FOR APP USERS

Emotions About Paper-based Survey Completion

Interacting with paper-based materials can induce different feelings. This survey refers to the emotions you may have experienced when using the paper-based version of these surveys. Specifically, the following questions pertain to feelings you experienced WHILE using the paper-based version of these surveys today. Record your answers below, using the appropriate number, where 1 indicates that you strongly disagree with a statement and 5 indicates that you strongly agree.

Questions:

I enjoyed using the paper-based version of these surveys.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Using the paper-based version of these surveys bored me.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Using the paper-based version of these surveys annoyed me.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Using the paper-based version of these surveys made me feel anxious.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

LIFE HISTORY QUESTIONNAIRE

1. What is your age? ___ years old (Please type your age in the blank provided)
2. What is your gender? (Please type an "X" in the blank next to one response)

   __ Female

   __ Male

   __ Other/Prefer not to say

3. Does anyone in your **biological family** have a history of **diagnosed** mental illness? (Please mark all that apply with an "X".)

   Father:

   __ None/Unknown

   __ Depression

   __ Anxiety/panic disorder

   __ Bipolar disorder

   __ Attention deficit hyperactivity disorder

   __ Alcohol/drug abuse

   __ Personality disorder

   __ Other

   Mother:

   __ None/Unknown

   __ Depression

   __ Anxiety/panic disorder

   __ Bipolar disorder

   __ Attention deficit hyperactivity disorder

   __ Alcohol/drug abuse

   __ Personality disorder

__ Other

Sibling(s):

__ None/Unknown

__ Depression

__ Anxiety/panic disorder

__ Bipolar disorder

__ Attention deficit hyperactivity disorder

__ Alcohol/drug abuse

__ Personality disorder

__ Other

Aunt(s)/Uncle(s):

__ None/Unknown

__ Depression

__ Anxiety/panic disorder

__ Bipolar disorder

__ Attention deficit hyperactivity disorder

__ Alcohol/drug abuse

__ Personality disorder

__ Other

Grandparent(s):

__ None/Unknown

__ Depression

__ Anxiety/panic disorder

__ Bipolar disorder

__ Attention deficit hyperactivity disorder

__ Alcohol/drug abuse

__ Personality disorder

__ Other

Child(ren):

__ None/Unknown

__ Depression

__ Anxiety/panic disorder

__ Bipolar disorder

__ Attention deficit hyperactivity disorder

__ Alcohol/drug abuse

__ Personality disorder

__ Other

4. Have you ever experienced a traumatic (i.e., severely distressing) event?

__ None

__ Physical assault

__ Sexual assault

__ Severe accident

__ Witnessed a tragedy

__ Other

5. At what age did you first feel mentally unwell? ___ years old

6. Did the onset of your symptoms coincide with a stressful life event? (Check all that apply.)

__ None

__ Death of loved one

__ Loss of employment or financial loss

__ Physical injury or medical diagnosis

__ Separation/divorce

__ Other

7. How long has it been since you last felt well? (Please enter an approximate number in one blank; e.g., “5 years” or “10 weeks” or “9 days”) ___ years; ___ weeks; ___ days

System Usability Scale (adapted for EarlyDetect mobile app)

On a scale of 1 to 5, with 1 = strongly disagree and 5 = strongly agree, please rate your agreement with the following statements.

		1 = Strongly disagree 5 = Strongly agree				
	Question	1	2	3	4	5
1	I think that I would like to use the EarlyDetect APP frequently.	☐	☐	☐	☐	☐
2	I found the EarlyDetect APP unnecessarily complex.	☐	☐	☐	☐	☐
3	I thought the EarlyDetect APP was easy to use.	☐	☐	☐	☐	☐
4	I think that I would need the support of a technical person to be able to use the EarlyDetect APP.	☐	☐	☐	☐	☐
5	I found the various functions in EarlyDetect APP were well integrated.	☐	☐	☐	☐	☐
6	I thought there was too much inconsistency in the EarlyDetect APP.	☐	☐	☐	☐	☐
7	I would imagine that most people would learn to use the EarlyDetect APP very quickly.	☐	☐	☐	☐	☐
8	I found the EarlyDetect APP very cumbersome to use.	☐	☐	☐	☐	☐
9	I felt very confident using the EarlyDetect APP.	☐	☐	☐	☐	☐
10	I needed to learn a lot of things before I could get going with the EarlyDetect APP.	☐	☐	☐	☐	☐

The SUS adapted from Brooke (1996) for the EarlyDetect mobile app

System Usability Scale (adapted for paper questionnaires)

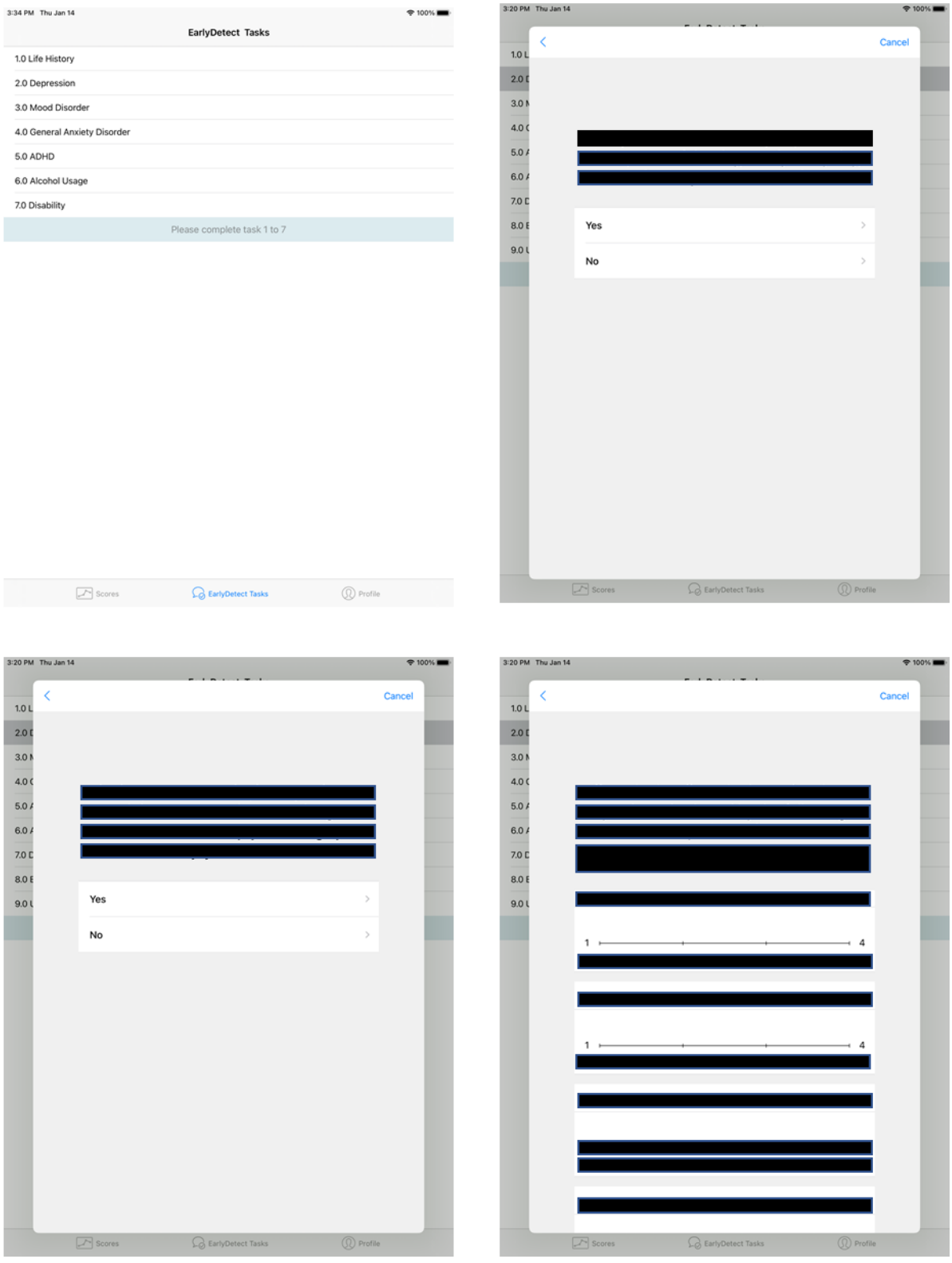
On a scale of 1 to 5, with 1 = strongly disagree and 5 = strongly agree, please rate your agreement with the following statements.

		1 = Strongly disagree 5 = Strongly agree				
	Question	1	2	3	4	5
1	I think that I would like to use the EarlyDetect questionnaire frequently.	☐	☐	☐	☐	☐
2	I found the EarlyDetect questionnaire unnecessarily complex.	☐	☐	☐	☐	☐
3	I thought the EarlyDetect questionnaire was easy to use.	☐	☐	☐	☐	☐
4	I think that I would need the support of a technical person to be able to use the EarlyDetect questionnaire.	☐	☐	☐	☐	☐
5	I found the various functions in the EarlyDetect questionnaire were well integrated.	☐	☐	☐	☐	☐
6	I thought there was too much inconsistency in the EarlyDetect questionnaire.	☐	☐	☐	☐	☐
7	I would imagine that most people would learn to use the EarlyDetect questionnaire very quickly.	☐	☐	☐	☐	☐
8	I found the the EarlyDetect questionnaire very cumbersome to use.	☐	☐	☐	☐	☐
9	I felt very confident using the the EarlyDetect questionnaire.	☐	☐	☐	☐	☐
10	I needed to learn a lot of things before I could get going with the EarlyDetect questionnaire.	☐	☐	☐	☐	☐

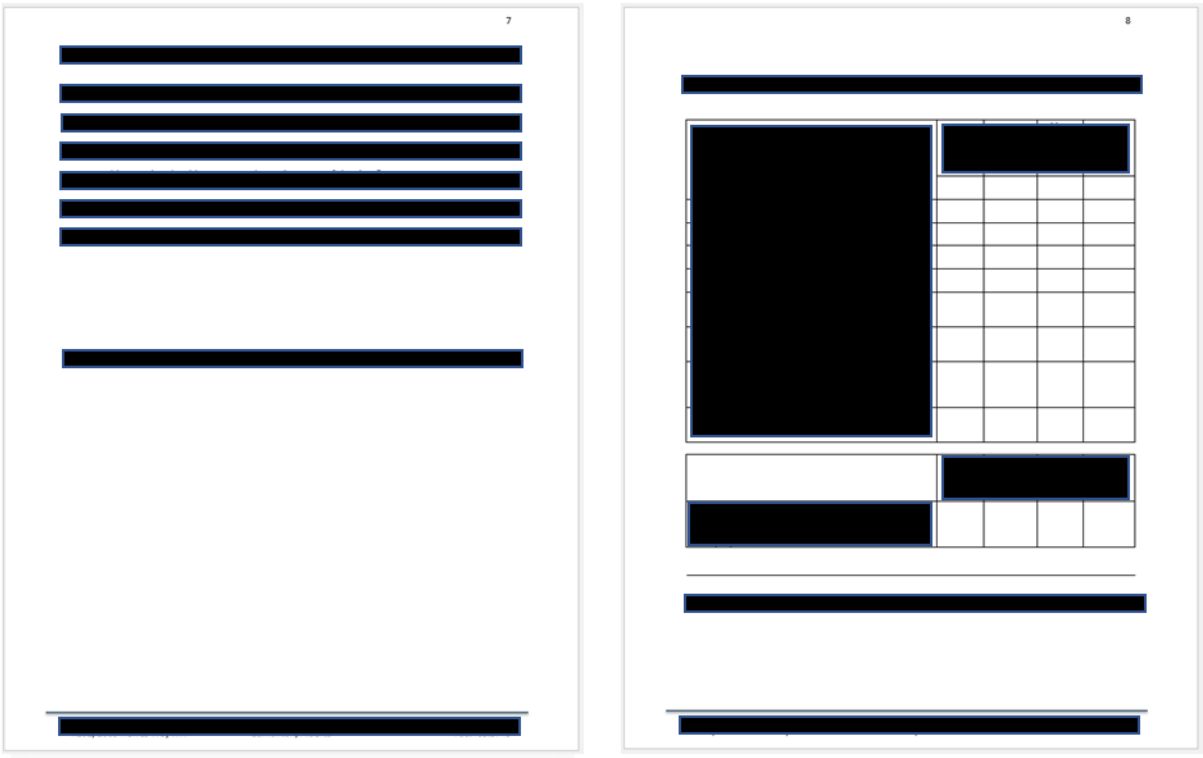
The SUS adapted from Brooke (1996) for the EarlyDetect paper-and-pencil questionnaires.

Figure S1

a) iPad screenshot



b) Paper version layout



Note. Selected contents were masked for copyright reasons. Example screenshots showing EarlyDetect's views: a) the navigation structure (top left), MINI Major Depressive Episode assessment (top right and bottom left) and PHQ-9 (bottom right) on an iPad, and b) the corresponding views on the paper version.