Accepted Manuscript

To appear at The 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 2022.

Can Robots Help in the Evaluation of Mental Wellbeing in Children? An Empirical Study

Nida Itrat Abbasi[1,*], Micol Spitale[1], Joanna Anderson[2], Tamsin Ford[2], Peter B. Jones[2] and Hatice Gunes[1]

***Abstract*— Socially Assistive Robots (SARs) show promise in helping children during therapeutic and clinical interventions. However, using SARs for the evaluation of mental wellbeing of children has not yet been explored. Thus, this paper presents an empirical study with 28 children 8-13 years old interacting with a Nao robot in a 45-minute session where the robot administered (robotised) the Short Mood and Feelings Questionnaire (SMFQ) and the Revised Child Anxiety and Depression Scale (RCADS). Prior to the experimental session, we also evaluated children's wellbeing using established standardised approaches via online RCADS questionnaires filled by the children (self-report) and their parents (parent-report). We clustered the participants into three groups (lower, medium, and higher tertile) based on their SMFQ scores. Further, we analysed the questionnaire responses across the three clusters and across the different modes of administration (self-report, parent-report, and robotised). Our results show that the robotised evaluation seems to be the most suitable mode in identifying wellbeing related anomalies in children across the three clusters of participants as compared with the self-report and the parent-report modes. Further, children with decreasing levels of wellbeing (lower, medium and higher tertiles) exhibit different response patterns: children of higher tertile are more negative in their responses to the robot while the ones of lower tertile are more positive in their responses to the robot. Findings from this work show that SARs can be a promising tool to potentially evaluate mental wellbeing related concerns in children.**

I. Introduction

The current COVID-19 pandemic has led to societal changes that have severely impacted the mental wellbeing especially of the vulnerable population groups like children - due to social isolation, study-from-home arrangements, limited finances, academic pressures and social distancing from peers and friends [1]. Even before the pandemic era, anxiety and depression had increased among children in the UK, resulting in a growing rate of suicide numbers [1]. This increase in the number of children with wellbeing related concerns has turned into a real call for action by the healthcare and psychological support fields due to the limited resources and struggle in understanding and addressing children's needs. Besides the current barriers to resources and accessibility, children are also very reluctant to disclose sensitive information to other adults and it often takes months of dedicated trust building for therapists to get the children to open up [2], [3].

Socially Assistive Robots (SARs) [4] are a promising avenue for overcoming those barriers and meeting the children's needs. Past works have already demonstrated the potential of SARs in supporting vulnerable populations (e.g., children with autism [5], elderly people [6]). For example, Ramirez et al. [7] presented a comprehensive robot-assisted intervention for children with autism showing the conditions in which a robot-based session can be useful to assess autism risk factors for diagnosing autism. Also, Spitale et al. [8] explored the use of physical conversational agent (i.e., a puppet robot) to support linguistic assessment and training among children with language impairment. However, to the best of our knowledge, no study has yet explored whether robots can be used to aid in the evaluation of children's mental wellbeing.

To this end, in this work we have undertaken an empirical study involving 28 children between the ages of 8-13 years old who interacted with a Nao robot during a 45 mins-long session. During the interaction, the robot performed four different tasks in the following order: i) open-ended questions about happy and sad memories over the last week, ii) administering the Short Mood and Feelings Questionnaire (SMFQ) [9], iii) administering the picture task inspired by the Children's Apperception Test (CAT) [10], where children are asked to answer questions related to pictures shown, and iv) administering the Revised Children's Anxiety and Depression Scale (RCADS) [11] for generalised anxiety, panic disorder and low mood (*robotised*). Prior to the experiment session, we have also evaluated children's wellbeing using the established standardised approaches via online questionnaires filled by children (*self-report*) and parents (*parent-report*) of the RCADS. We first clustered the participants into three groups (lower, medium, and higher tertile) based on their SMFQ scores (administered by the robot during the experiment session) and then investigated which mode would be most effective in identifying children's cases with wellbeing related concerns. We have also investigated whether the clusters have different response patterns across the three modes of test administration (self-report, parent-report and robotised).

The main findings of this study are as follows: (1) robot-administered tests (robotised) seem to be more successful in identifying cases in children with wellbeing related anomalies as compared with other established standardised modes of test administration (self-report and parent-report); and (2) for children belonging to the higher tertile, the robot has elicited more negative response ratings as compared with the standardised modes (self-report and parent-report); on the other hand, for children belonging to the lower tertile, the robotised mode of administration has elicited more positive

[1] Department of Computer Science and Technology, University of Cambridge, UK. *Corresponding author: nia22@cam.ac.uk [2] Department of Psychiatry, University of Cambridge, UK.

response ratings as compared with the standardized modes (self-report and parent-report).

II. Related Works

A. Children's wellbeing assessment

Evaluation of children's wellbeing is an integral part of understanding children's welfare and building initiatives that promote overall mental health in children [12]. In the UK alone, the annual expenditure on children and young people's mental health has been estimated at 79 million pounds[1]. Several surveys like, the Mental Health of Children and Young People Surveys (MYHCP, conducted in the UK in 1999, 2004, 2017) [13], the Oxwell survey for school mental health (conducted in the UK in 2019, 2020, 2021) [14], the Australian Child and Adolescent Surveys of Mental Health and Wellbeing (conducted in Australia between 1998 and 2000) [15], the Young Minds Matter (conducted in Australia between 2013 and 2014) [16] and many more, employing the traditional paper-pen response or online self-reporting of answers to psychometric questionnaires, in order to understand mental wellbeing of children. While these surveys have been effective in understanding the prevalence of mental health related concerns in children, they are heavily based on the assumption that the self-report answers are 'true' or 'correct' [17]. However, some children might be reluctant to disclose personal and sensitive information about themselves [3], which could result in incorrect responses to the administered questionnaires. Further, in the case of human-administered psychological assessments, children might be intimidated by the presence of another adult whom they might consider as 'being in-charge', thus, ending up providing expected answers rather than reporting about their true experiences and feelings [17]–[19].

B. CHRI in healthcare and wellbeing

Child robot interaction (CHRI) has been very effective in "gaining trust" of the children and has been used as an effective tool for providing support in health [20] and education settings [21], as well as providing companionship, for instance, to make the hospital visits less intimidating for children [22]. Several health related initiatives - the Adaptive Strategies for sustainable long-term interaction (ALIZ-E) project for developing robot companions for better management of diabetes in children (7-11 years old) [20], the Development of Robot-Enhanced therapy for children with Autism spectrum disorders (DREAM) project for developing autonomous social robot therapy to cater to the behavioural traits for children with ASD [23] - have successfully used robots for providing guidance and companionship to better understand the conditions of these children. Robots have also been used in schools for tutoring and providing academic and non-academic help [21]. Moreover, the acceptance of robot teaching aids have been observed across all age groups (primary school [24], 3-5 years [25] and above 12 years [26]).

[1]https://www.gov.uk/government/news/79-million-to-boost-mental-health-support-for-children-and-young-people

Robots have been proven to be very useful in making children divulge confidential and private information. For example, robots have been successfully used to identify occurrences of bullying, violence and abuse in children [19], [27]. Robots have also been used to assess risk factors for autism in children [7] and also for linguistic evaluation and training in children with language impairments [8]. In many cases, robots have not only been able to extract sensitive knowledge from children [28] but also influenced their opinions on previously occurring transgressions [29]. However, robots have not yet been explored as tools for the pro-active evaluation of mental wellbeing in children. Moreover, how the response patterns of children with varying wellbeing related concerns change when interacting with a robot as compared to other established approaches of wellbeing measures (self-report or parent-report) has also not been investigated. Thus, we address the above points through this empirical study of 8-13 years old children.

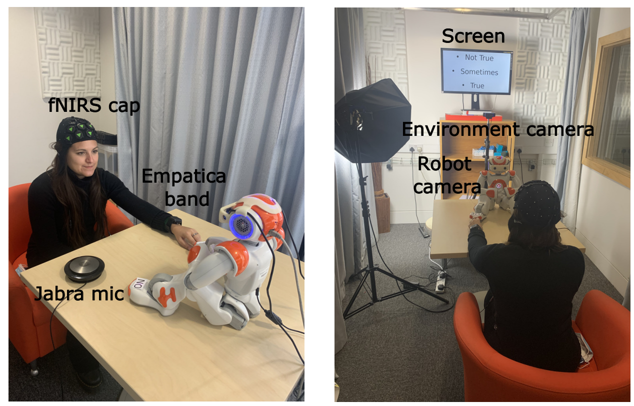

Fig. 1. Figures depicting the experimental setup that includes: the Nao robot, fNIRS cap, Empathica wrist band, Jabra mic, the Nao robot, the external camera, and the screen. The left figure depicts the lateral view of the setup where the participant is seated in front of the Nao robot, the right figure shows the perspective of the same setup from the back. Actual pictures from the CHRI sessions have not been used in order to protect the privacy of the child.

III. Research questions

In this study, our main goal was to understand to what extent a humanoid robot can aid in the evaluation of children's wellbeing, and whether and how this would differ from the self-report and the parent-report responses. Thus, we have investigated which modes of test administration (self-report, parent-report, and robotised) would be more beneficial in identifying mental wellbeing related anomalies (**RQ1**).

Previous works have also shown that children with depression and anxiety have lowered response initiation, decreased assertiveness and increased competitiveness [30], [31], leading to varying response patterns in electronic games [31], internet use, other cyber behaviour [32] and even basic tasks like drawing [33] and reading [34]. However, it has not been investigated whether those children have different response patterns to robotised tests as compared with the established standards (self-report and parent-report). Therefore, in this study we have also explored how children with varying levels of mental wellbeing (lower, medium and higher tertiles) respond to the different modes of test administration (self-report, parent-report, and robotised) **(RQ2)**.

IV. The Study

To find answers to these two research questions, we conducted an experimental study involving 28 children aged between 8-13 years old. Before the experiment, each child and their parent completed the Revised Child Anxiety and Depression Scale based on the children's wellbeing (RCADS self-report and RCADS parent-report). The time between the online form filling and the experiment session ranged from less than a few hours to more than a few weeks. After that, we invited the participants into our lab to interact with a Nao robot in a 45 mins session, in which the robot undertook four tasks with the children that aimed to aid in the evaluation of the child's wellbeing. Figure 1 depicts the experimental setup of the study where a participant (in this picture, an experimenter who acted as one of the participants to protect the privacy of the participants) interacts with the Nao robot.

A. Tasks

The robot asked children to perform the following four tasks that are established and recognised as a standard in child cognitive psychology and HRI literature [19], [35]–[38]. The children were asked to respond by verbalising their answers.

Task 1. The robot asked open-ended questions to the child about a recent happy memory and about a recent negative memory. This task aims to identify any surface psychological distress that the child may be experiencing [37], [39].

Task 2. The robot administered the Short Mood and Feelings Questionnaire (SMFQ) [9]. The robot made statements from the SMFQ (e.g., "You didn't enjoy anything at all") while the child answered by verbalising his/her responses (i.e., the child could pick one option among "True", "Sometimes", "Not true", and say it to the robot). The experiment room contained an external screen that acted as a visual aid for the children by displaying the response options, so they did not have to memorize them (as shown in Figure 1).

Task 3. The robot performed a picture task inspired by the Children's Apperception Test (CAT) [10], which consists of showing a series of images (we showed picture 7, picture 9 and picture 10 from CAT since they are more relevant with our research theme) on a screen to the children and asking them questions like "What do they see in the picture?", "What do they think happened before the picture?" and "What do they think happened after the picture?". This task aimed to assess the personality and wellbeing related traits of the children, and variations of this task have also been previously used in HRI studies [35]).

Task 4. The robot administered the Revised child Anxiety and Depression (RCADS) questionnaires [11] (described in Section V-A). For our experiment session, we have only included the subscales corresponding to Genaralised Anxiety (6 items), Panic Disorder (9 items), and Low Mood (10 items) as they were the most relevant for our work. The items of the questionnaires are in statement format (e.g., "You worry about things", "When you have a problem, your heart beats really fast"), and the children have been asked to respond by verbalising their response by choosing one among the ones displayed on the screen ("Never", "Sometimes", "Often", "Always") (as shown in Figure 1). Since in this work, we aim to compare the modes of psychological test administration with respect to the evaluation of children's wellbeing, we have only considered quantitative measures (Task 2 SMFQ and Task 4 RCADS) for further analysis in this work.

B. Robotic platform

The experiment was conducted using the Nao humanoid robotic platform that is equipped with state-of-the-art interactive features like vision sensors for object detection, human-like movement and speech production. We defined the robot's level of autonomy based on the framework in [40] as follows: sense (not autonomous), plan (semi-autonomous, using pre-scripted decisions based on children's behavior), and act (fully autonomous). Nao followed a pre-written script while administering the experiment and the robot movements have been also pre-programmed. The robot also engaged with the participant through head and arm movements, and by "looking" at the participant during the interaction.

C. Participants

Participation in the study was on a voluntary basis. We followed two strategies for subject recruitment: online advertisement circulated within the local schools and snowball sampling among contacts within the study team. We recruited 28 children from the age groups of 8-13 years old (21 females and 7 males, mean age= 9.5 years old, SD = 1.5 years old) within Cambridgeshire in the United Kingdom. The recruitment criteria included exclusion on the basis of any current neurological and psychological disorders. For this purpose, only the parents that declared that their children do not have any psychological and neurological issues were considered for this study. Interested volunteers could respond to the expression-of-interest online form, following which they were contacted to complete the online questionnaires (RCADS self-report and RCADS parent-report) prior to their experimental session. Upon completion of the online questionnaires, their participation was confirmed and they were notified about the date and the time of the experiment session based on their availability.

D. Protocol

The experiment took place in a dedicated room where each child interacted with the Nao robot in a dyadic interaction setting as depicted in Figure 1. The room was equipped with a mirror glass that allowed the experimenters and the parents to monitor the whole session without being seen by the children. Each child was requested to sit on a chair that was placed 1.5m away from the robot positioned on a table in front of the child. We also placed an external screen behind the robot - to visually display response ratings or pictures used in the study. We audio and video recorded the participants' session using Jabra microphone, robot and external cameras. Moreover, physiological data (heart rate and wrist movements using Empatica band) and brain activity (functional Near-Infrared Spectroscopy (fNIRS) using Lumo device) were also collected during the experiment session (Figure 1). Once the child was seated in the designated position and the equipment was placed, the experimenters left the room to monitor the interaction from the control room. The interaction started with the Nao robot introducing itself, giving salutations, offering to fist bump or wiping his forehead actions for the child. The robot also asked introductory questions like "how was your day?" for ice-breaking and making the child more comfortable in the experimental session [19], [38]. This enabled the child to get familiar with the robot and its functionality. After the warm-up session, the child performed sequentially the four tasks described in Section IV-A. For open-ended tasks (Task 1: memory recall and Task 3: picture task), the child was asked to answer the robot's question verbally, and then press the toe button of the Nao robot when they were done speaking. For other tasks like the SMFQ and the RCADS, the robot would continue with the subsequent statement in case of incomplete or no response after a pre-determined time ($5s$) for that task has lapsed [37]. Besides Task 1, all the other tasks prompted the child with some text or image displayed on the screen placed behind the robot. For example, in Task 4, the screen displayed the answers (i.e., "Never", "Sometimes", "Often", "Always") of the RCADS questionnaire in order to help the child with the response ratings. Regular breaks were provided after the end of each task, and if the child wanted to take a break, the experimenters entered the room and let them have a break. The children were also given an option to speak with their accompanying guardians and/or drink water. The children were also instructed that they may stop the interaction at any time or skip certain parts of the interaction depending on how they felt. The experiment protocol was approved by the relevant ethics committees of the University of Cambridge (PRE.2021.036).

V. Data Analysis

A. Measures

Children have been asked to complete four tasks that have been inspired by standardized tests [35], [36]. Since in this work we aim to compare the children's responses between the three modes of test administration (self-report vs. parent-report vs. robotised), tasks that contained standardised tests in psychology literature (Task 2: SMFQ and Task 4: RCADS) were considered for analysis in this work. For both Task 2 and Task 4, we computed the children's aggregated scores following the scoring methodology for the questionnaires (SFMQ [9] (0 - "Not True", 1 - "Sometimes", 2 - "True") and RCADS [11] (0-"Never", 1-"Sometimes", 2-"Often", 3-"Always"). We computed the following variables:

1. Total score from SFMQ, as the sum of the item answers of the questionnaires using the pre-established scoring (0 - "Not True", 1 - "Sometimes", 2 - "True").

2. Generalised anxiety subscale from the RCADS as the sum of the first 6 items of the questionnaires using the questionnaires scoring (0-"Never", 1-"Sometimes", 2-"Often", 3-"Always").

3. Panic disorder subscale from the RCADS as the sum of the next 9 items of the questionnaires using the questionnaires scoring.

4. Low mood subscale from the RCADS as the sum of the next 10 items of the questionnaires using the questionnaires scoring.

B. Data Analysis

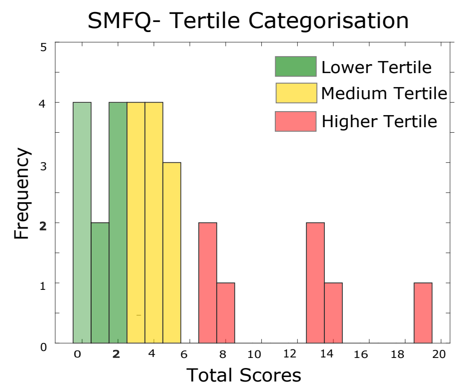

Fig. 2. Tertile categorisation of the SMFQ scores to create participant clusters: lower tertile (score <=2), medium tertile (2<score<=5) and higher tertile (score >5).

1) Questionnaires ratings: For the robotised measure, the audio files collected during the experiment session were divided into four tasks using the Audacity software [41]. Then, the files were transcribed using the DeepSpeech library[2] while the errors in transcription were tweaked manually by the experimenters. For the self-report and parent-report tests, Qualtrics[3] online links were sent to the parents prior to the experiment and the responses were recorded as CSV files.

2) Statistical analysis: To analyze the data collected, we divided participants into three clusters (tertiles) based on the total scores computed from the SMFQ (during the experiment session with the robot) corresponding to the lower tertile, medium tertile and higher tertile (see Figure 2) similar to the categorisation followed by [42]. Tertile edges for the total scores of participants resulted in 10 participants in the lower tertile category (SMFQ scores $<=2$), 11 participants in the medium tertile category ($2<$ SMFQ scores $<=5$) and 7 participants in the higher tertile category (SMFQ scores > 5). Since SMFQ provides a clinical cut-point of depression in children, the lower tertile is highly unlikely to have a diagnosis of depression, as are the medium tertile, while some in the highest tertile possibly or probably have a diagnosis of depression. The SMFQ data was found to not follow a normal distribution (using the Kolmogorov–Smirnov

[2]https://github.com/mozilla/DeepSpeech
[3]https://www.qualtrics.com/

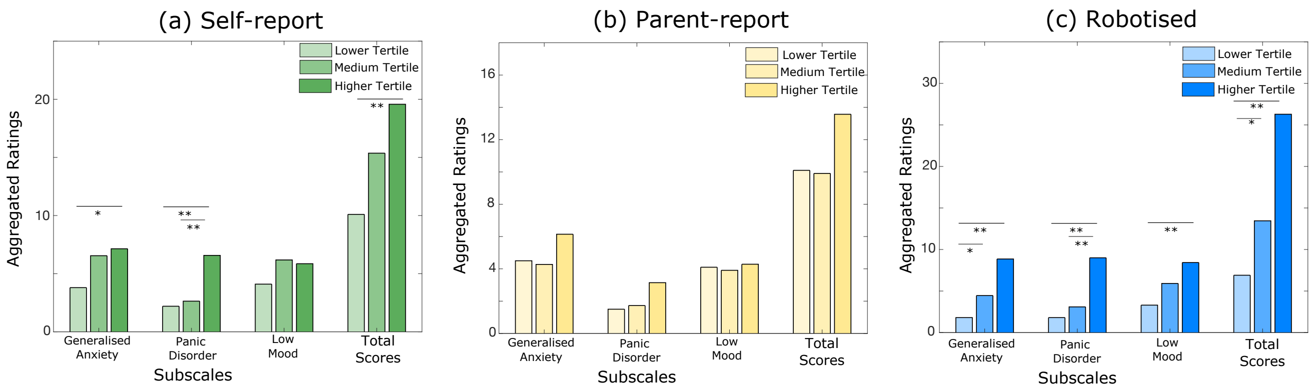

Fig. 3. Mean values of the RCADS subscale scores and the total scores for (a) self-report, (b) parent-report, and (c) robotised modes of questionnaire administration. *p< 0.05 uncorrected, **p< 0.05 corrected.

test) and thus we used non-parametric statistical tests. We conducted two-sided Wilcoxon Rank Sum test to investigate the difference in response ratings between the various modes of test administration (self-report, parent-report and robotised) across the varying levels of wellbeing (lower tertile, medium tertile and higher tertile) to address RQ1. Then, we used two-sided Wilcoxon Rank Sum tests to compare the difference in response ratings for children with varying levels of wellbeing (lower tertile, medium tertile and higher tertile) across the parent-report and robotised modes of test administration to answer RQ2. We also conducted two-sided Wilcoxon Signed Rank tests to compare between the varying levels of wellbeing (lower tertile, medium tertile and higher tertile) between the self-report and the robotised modes of test administration to also answer RQ2. False discovery rate was used to correct for all the computed p values from Type 1 error.

VI. RESULTS

A. Difference in the participant clusters across different modes

We ran Wilcoxon rank-sum tests to compare the RCADS subscales (generalized anxiety, panic disorder, low mood, and total scores) of each mode (self-report, parent-report, and robotised) between the population groups (lower, medium, and high tertiles).

For the self-report version of the test (see Figure 3(a)), we found that the panic disorder subscale was significantly ($W = 79.5$, $p = 0.043$ corrected) higher in the higher tertile (Median $(Mdn) = 7$) than in the medium tertile ($Mdn = 2$) of the population. Our results also showed that the panic disorder subscale was statistically significantly ($W = 66$, $p = 0.041$ corrected) higher in the higher tertile ($Mdn = 7$) as compared with the lower tertile ($Mdn = 2.5$). While considering the total scores, our results showed that the higher tertile ($Mdn = 22$ was statistically significantly ($W = 66.5$, $p = 0.041$ corrected) higher than the lower tertile ($Mdn = 10$). For the parent-report version of the RCADS, we did not find any comparison to be statitstically significant as depicted in Figure 3(b).

For the robotised tests, the results showed that the generalised anxiety subscale was statistically significantly ($W = 63$, $p = 0.031$ corrected) lower in the lower tertile ($Mdn = 1$) than in the higher tertile ($Mdn = 10$) groups as shown in Figure 3(c). Statistically significant changes were also found in the panic disorder subscale. The panic disorder was significantly ($W = 58.5$, $p = 0.016$ corrected) higher in the higher tertile ($Mdn = 10$) than in the lower tertile ($Mdn = 2$) group. Moreover, the panic disorder subscale was significantly ($W = 72.5$, $p = 0.016$ corrected) higher in the higher tertile ($Mdn = 10$) than in the medium tertile ($Mdn = 2$) group. For the low mood subscale, the results showed that the response ratings in the higher tertile ($Mdn = 8$) are statistically significantly ($W = 66.5$, $p = 0.041$ corrected) higher as compared to the response ratings in the lower tertile ($Mdn = 3$). Finally, as shown Figure 3(c), the total scores of the three reported subscales of RCADS was significantly ($W = 60$, $p = 0.016$ corrected) higher in the higher tertile ($Mdn = 30$) than in the lower tertile ($Mdn = 5$). Out of three modes of test administration, the robotised mode seems to highlight more the differences in cases of wellbeing related anomalies, followed by the self-report mode, and then the parent-report mode.

B. Difference in modes of administration across the clusters

We conducted Wilcoxon rank-sum tests to compare the RCADS subscales (generalized anxiety, panic disorder, low mood, and total scores) of each population group (lower, medium, and higher tertiles) between parent-report and robotised modes of test administration. We have also conducted the Wilcoxen signed-rank tests to compare the RCADS subscales (generalized anxiety, panic disorder, low mood, and total score) of each population group (lower, medium, and higher tertiles) between self-report and robotised modes of test administration.

For the lower tertile group (see Figure 4 (a)), a Wilcoxon rank-sum test indicated that the generalised anxiety subscale was statistically significantly ($W = 136$, $p = 0.041$ corrected) higher in the parent-report mode of the test administration($Mdn = 5$) than in the robotised test ($Mdn =$

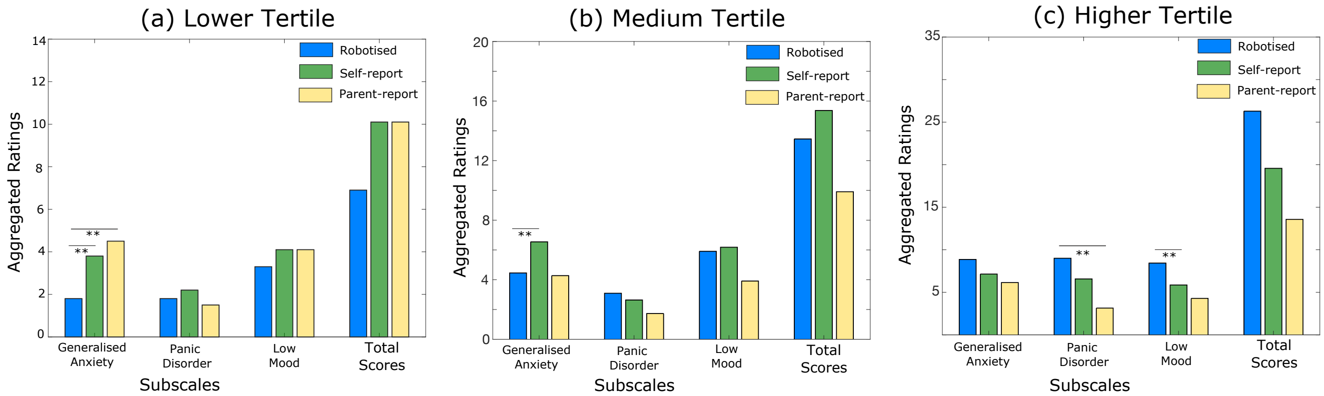

Fig. 4. Mean values of the RCADS subscales and total scores for (a) lower tertile, (b) medium tertile, and (c) higher tertile participant clusters. *p$<$ 0.05 uncorrected. **p$<$ 0.05 corrected.

1), as shown in the Figure 4(a). We have also found that the generalised anxiety subscale was statistically significantly ($W = 0$, $p = 0.041$) higher for the self-report ($Mdn = 4$) mode of test administration as compared with the robotised test ($Mdn = 1$) as seen in the Figure 4(a).

For the medium tertile population group, a Wilcoxon sign rank test indicated that generalised anxiety subscale was significantly ($W = 0, p = 0.026$ corrected) higher in the self-report ($Mdn = 6$) versions of the test than in the robotised ($Mdn = 4$) version, as shown in Figure 4(b).

For the higher tertile population groups, the Wilcoxon rank sum test indicated that the panic disorder subscale was significantly ($W = 35$, $p = 0.041$ corrected) higher in the robotised test ($Mdn = 10$) than in the parent-report ($Mdn = 2$) test as depicted in Figure 4(c). We have also observed that for the low mood subscale, the robotised response ratings were statistically significantly higher ($W = 28$, $p = 0.041$) in the robotised test ($Mdn = 8$) as compared with the self-report ($Mdn = 7$) mode of test administration (see Figure 4(c)).

To sum up, our results showed that, in the robotised administered mode, the generalized anxiety, panic disorder, low mood subscales and total score were significantly higher for the higher tertile group than for the lower tertile group. We also found that the self-report administered test showed the panic disorder to be significantly higher in the higher tertile than in the medium and lower tertile groups. Also, we observed that children of the higher tertile group seemed to respond more negatively in the robotised mode in comparison with the other two modes (self-report and parent-report) of test administration.

VII. Discussion

A. What modes of administration would be more beneficial in identifying cases of mental wellbeing anomalies? (RQ1)

Our results showed that out of the three administration modes, robotised tests appeared to amplify the differences between children with varying levels of mental wellbeing concerns (lower tertile, medium tertile and higher tertile) as shown in Figure 3. The subscales and the total scores of RCADS for the robotised mode are significantly higher in children belonging to the higher tertile than those belonging to the lower tertile (see Figure 3(c)). These findings are in line with prior research, where robotic procedures for psychometric assessment have shown great potential in the detection of mild cognitive impairment in the elderly [43] and healthy adults [37]. Our study extends these findings to children and suggests that robots may be successfully used to distinguish between children experiencing mental wellbeing related problems as compared to children that are not.

Another significant aspect is to understand why the robotised mode of test administration seems to be more beneficial in detecting mental wellbeing related anomalies in children. Prior research has suggested that children consider robots as social companions [44] and peers [45], [46]. Moreover, robots have also been shown to improve communication between children suffering from anxiety, their parents and their therapists [47]. Thus, robots could offer a more unique and comforting way - among other different reasons - to encourage children to open up that might not be possible in the traditional modes of psychological test administration. However, it is important to note that the role of the robots is to aid in the evaluation procedures and they are not meant to substitute the diagnosis and treatment provided by the clinical professionals.

B. How do children with varying levels of wellbeing respond to the different modes of administration? (RQ2)

We found that children belonging to the higher tertile group (Figure 4(c)) seemed to consistently respond more negatively in the robotised mode in comparison with the other two modes (self-report and parent-report) of the test administration. For this population, we observed that the robotised score had been consistently higher in response ratings followed by the self-report responses and then the parent-report responses, across all the subscales. Bethel et al. [28] have previously shown that children are less reluctant to divulge secret information to a human rather than a robot. More specifically, as seen from our study, children that are experiencing wellbeing related concerns (belonging to the higher tertile) respond in a more negative manner in front of the robot as compared to any other modes of test administration. It might be possible that children belonging to this category have found a friend or a confidante in the robot, leading them to answer more truthfully (disclosing sensitive information) about their feelings as prior research does show that children consider robots as social peers [45], [46]. While considering the parent-report scores for this group of children, the response ratings are observed to be more optimistic as compared with the other two modes (robotised and self-report). This finding is also consistent with previous psychological studies [48], [49] that showed that parents are often unable to recognise whether their children are experiencing mental wellbeing problems, leading them to seek psychological help for their children when it is very late.

For the lower tertile category as shown in Figure 4(a), predominantly an opposing trend has been observed, where the robotised ratings are the lowest, followed by the self-report ratings and then the parent-report ratings (except in case of panic disorder). Since these children might not be experiencing any mental wellbeing related problems, their experience with the robot might be proving to be very exciting and novel, eliciting more positive responses. Lenh et al. [50] have demonstrated a similar phenomenon of over-engagement in an edutainment setting, where classroom learning was not improved by the use of a Tangible User Interface (TUI) as compared with the traditional paper and pen method. Past works [21], [51] have shown that children enjoy performing tasks with the robots, and generally like their company. Thus, for children that are not experiencing mental wellbeing related problems (lower tertile), the CHRI experience might be proving to be too stimulating and engaging, eliciting more positive responses as compared with the self-report and parent-report modes of test administration.

For the medium tertile category, the observed trend seems to be in between the lower tertile and the higher tertile categories. Their ratings are predominantly highest for the self-report category followed by the robotised category and then the parent-report category.

VIII. Conclusions & Future Work

In this work, we conducted an empirical study to investigate the use of robots in the evaluation of children's mental wellbeing as compared with the established standardised modes of psychometric test administration. We found that robots can be used as effective tools for identifying children that are experiencing mental wellbeing related problems. Moreover, we have also found that children with varying levels of wellbeing concerns interact differently with the robot. For children that might not be experiencing mental wellbeing related problems, we discovered that robotised measures elicited more positive response ratings. While for children that might be experiencing wellbeing related concerns, robots might have enabled them to divulge their true feelings and experiences, eliciting more negative response ratings.

Although our work addressed the research questions defined, it has several limitations. Firstly, we have only considered questionnaire and verbal responses, while a more comprehensive investigation should be undertaken by considering also the non-verbal aspects such as physiology, audio, and face and body gestures. Moreover, this study is a feasibility study on using robots for wellbeing evaluation in children, and we have not taken into account the validity of the mode of administration (i.e., comparison with clinician-administered tests). Other limitations include an imbalanced participant group (many more girls than boys), time lapsed between online questionnaire filling and the experiment session (since mental health fluctuates), bias due to the mode of administration (robotised) used for the SMFQ clustering, and the novelty effect [52] that could confound the responses of the participants.

In our future studies, we aim to address these limitations to understand more comprehensively the effectiveness of using robots for helping in the evaluation of mental wellbeing in children.

Data Access Statement

Overall statistical analysis of research data underpinning this publication is available in the text of this publication. Additional raw data related to this publication cannot be openly released; the raw data contains transcripts of interviews, but none of the interviewees consented to data sharing.

Acknowledgments

This work was supported by the University of Cambridge's Human Machine Collaboration (OHMC) Small Equipment Funding. N. I. Abbasi is supported by the W.D. Armstrong Trust PhD Studentship and the Cambridge Trusts. M. Spitale and H. Gunes are supported by the EPSRC project ARoEQ under grant ref. EP/R030782/1. All research at the Department of Psychiatry in the University of Cambridge is supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014, particularly T. Ford) and NIHR Applied Research Centre (P. Jones, J. Anderson). J. Anderson, T. Ford and P. Jones acknowledge support from the NIHR ARC East of England, and by the Cambridge BRC. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

References

[1] T. Ford and et al., "Mental health of children and young people during pandemic," 2021.

[2] P. Bharadwaj and et al., "Mental health stigma," *Economics Letters*, vol. 159, pp. 57–60, 2017.

[3] B. A. Pescosolido and et al., "Public knowledge and assessment of child mental health problems: Findings from the national stigma study-children," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 47, no. 3, pp. 339–349, 2008.

[4] A. Tapus and et al., "The grand challenges in socially assistive robotics," *IEEE Robotics and Automation Magazine*, vol. 14, no. 1, pp. N–A, 2007.

[5] A. Alabdulkareem and et al., "A systematic review of research on robot-assisted therapy for children with autism," *Sensors*, vol. 22, no. 3, p. 944, 2022.

[6] J. Abdi and et al., "Scoping review on the use of socially assistive robot technology in elderly care," *BMJ open*, vol. 8, no. 2, p. e018815, 2018.

[7] A. A. Ramírez-Duque and et al., "Robot-assisted intervention for children with special needs: A comparative assessment for autism screening," *Robotics and Autonomous Systems*, vol. 127, p. 103484, 2020.

[8] M. Spitale and et al., """ whom would you like to talk with?" exploring conversational agents for children's linguistic assessment," in *Proc. of the IDC Conference*, 2020, pp. 262–272.

[9] C. Sharp and et al., "The short mood and feelings questionnaire (smfq): a unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7-through 11-year-old children," *Journal of abnormal child psychology*, vol. 34, no. 3, pp. 365–377, 2006.

[10] L. Bellak and S. S. Bellak, "Children's apperception test." 1949.

[11] B. F. Chorpita and et al., "Revised children's anxiety and depression scale," *Hämtad från https://www. childfirst. ucla. edu/wp content/uploads/sites/163/2018/03/RCADSUsersGuide20150701. pdf*, 2015.

[12] K. D. Rosanbalm and et al., "Child wellbeing assessment in child welfare: A review of four measures," *Children and youth services review*, vol. 68, pp. 1–16, 2016.

[13] T. Ford and et al., "Data resource profile: the mental health of children and young people surveys (mhcyp)," *Int. journal of epidemiology*, 2020.

[14] K. L. Mansfield and et al., "Study protocol: the oxwell school survey investigating social, emotional and behavioural factors associated with mental health and well-being," *BMJ open*, vol. 11, no. 12, p. e052717, 2021.

[15] D. Lawrence and et al., "The mental health of children and adolescents: Report on the second australian child and adolescent survey of mental health and wellbeing," 2015.

[16] J. Hafekost and et al., "Introducing 'young minds matter'," *Australian Economic Review*, vol. 49, no. 4, pp. 503–514, 2016.

[17] R. Tourangeau and et al., "The psychology of survey response," 2000.

[18] M. K. Nock and A. E. Kazdin, "Parent expectancies for child therapy: Assessment and relation to participation in treatment," *Journal of Child and Family Studies*, vol. 10, no. 2, pp. 155–180, 2001.

[19] D. Godoi and et al., "Proteger: A social robotics system to support child psychological evaluation," in *2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE)*. IEEE, 2020, pp. 1–6.

[20] T. Belpaeme and et al., "Multimodal child-robot interaction: Building social bonds," *Journal of Human-Robot Interaction*, vol. 1, no. 2, 2012.

[21] H. Chen and et al., "Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement," *Computers & Education*, vol. 150, p. 103836, 2020.

[22] J. De Greeff and et al., "Activity switching in child-robot interaction: a hospital case study," in *ICSR*, vol. 8239. SPRINGER-VERLAG BERLIN, 2013, pp. 585–586.

[23] P. Gómez Esteban and et al., "A multilayer reactive system for robots interacting with children with autism," *arXiv e-prints*, pp. arXiv–1606, 2016.

[24] P. Baxter and et al., "Robot education peers in a situated primary school study: Personalisation promotes child learning," *PloS one*, vol. 12, no. 5, p. e0178126, 2017.

[25] G. Venture and et al., "Dance with me! child-robot interaction in the wild," in *ICSR*. Springer, 2017, pp. 375–382.

[26] A. Causo and et al., "Design of robots used as education companion and tutor," in *Robotics and mechatronics*. Springer, 2016, pp. 75–84.

[27] C. L. Bethel and et al., "Using robots to interview children about bullying: Lessons learned from an exploratory study," in *IEEE RO-MAN 2016*. IEEE, 2016, pp. 712–717.

[28] C. Bethel and et al., "Secret-sharing: Interactions between a child, robot, and adult," in *2011 IEEE Int. Conf. on systems, man, and cybernetics*. IEEE, 2011, pp. 2489–2494.

[29] J. Peter and et al., "Can social robots affect children's prosocial behavior? an experimental study on prosocial robot models," *Computers in Human Behavior*, vol. 120, p. 106712, 2021.

[30] A. E. Kazdin and et al., "Assessment of overt behavior and childhood depression among psychiatrically disturbed children." *Journal of Consulting and Clinical Psychology*, vol. 53, no. 2, p. 201, 1985.

[31] C. M. Rockhill and et al., "Friendship interactions in children with and without depressive symptoms: Observation of emotion during game-playing interactions and post-game evaluations," *Journal of abnormal child psychology*, vol. 35, no. 3, pp. 429–441, 2007.

[32] W. Rikkers and et al., "Internet use and electronic gaming by children and adolescents with emotional and behavioural problems in australia–results from the second child and adolescent survey of mental health and wellbeing," *BMC public health*, vol. 16, no. 1, pp. 1–16, 2016.

[33] R. Silver, "Identifying children and adolescents with depression: Review of the stimulus drawing task and draw a story research," *Art Therapy*, vol. 26, no. 4, pp. 174–180, 2009.

[34] A. E. Grills-Taquechel and et al., "Anxiety and reading difficulties in early elementary school: evidence for unidirectional-or bi-directional relations?" *Child Psychiatry & Human Development*, vol. 43, no. 1, pp. 35–47, 2012.

[35] P. Bremner and et al., "Personality perception of robot avatar teleoperators," in *ACM/IEEE HRI 2016*. IEEE, 2016, pp. 141–148.

[36] B. Cliffe and et al., "Digital cognitive behavioral therapy for insomnia for adolescents with mental health problems: feasibility open trial," *JMIR mental health*, vol. 7, no. 3, p. e14842, 2020.

[37] A. Di Nuovo and et al., "Assessment of cognitive skills via human-robot interaction and cloud computing," *Journal of bionic engineering*, vol. 16, no. 3, pp. 526–539, 2019.

[38] K. L. H. Ting and et al., "Integrating the users in the design of a robot for making comprehensive geriatric assessments (cga) to elderly people in care centers," in *IEEE RO-MAN 2017*. IEEE, 2017, pp. 483–488.

[39] K. A. Arditte Hall and et al., "Positive memory enhancement training for individuals with major depressive disorder," *Cognitive Behaviour Therapy*, vol. 47, no. 2, pp. 155–168, 2018.

[40] J. M. Beer and et al., "Toward a framework for levels of robot autonomy in human-robot interaction," *Journal of human-robot interaction*, vol. 3, no. 2, p. 74, 2014.

[41] T. Audacity, "Audacity," 2014.

[42] F. Munir and et al., "Occupational sitting time and its association with work engagement and job demand-control," in *European Academy of Occupational Health Psychology Conference*. European Academy of Occupational Health Psychology, 2014.

[43] S. Varrasi and et al., "A social robot for cognitive assessment," in *Companion of the 2018 ACM/IEEE HRI*, 2018, pp. 269–270.

[44] M. Nalin and et al., "Robotic companion for diabetic children: emotional and educational support to diabetic children, through an interactive robot," in *Proc. of the 11th int. conf. on interaction design and children*, 2012, pp. 260–263.

[45] D. Kumar Singh and et al., "Toy, tutor, peer, or pet? preliminary findings from child-robot interactions in a community school," in *Companion of the 2020 ACM/IEEE HRI*, 2020, pp. 325–327.

[46] J. de Greeff and et al., "Child-robot interaction in the wild: Field testing activities of the aliz-e project," in *ACM/IEEE HRI 2014*. IEEE, 2014, pp. 148–149.

[47] L. Arnold, "Emobie™: A robot companion for children with anxiety," in *ACM]/IEEE HRI 2016*. IEEE, 2016, pp. 413–414.

[48] J. Kim and et al., "Parent-child discrepancies in reporting of child depression in ethnic groups," *The Journal for Nurse Practitioners*, vol. 12, no. 6, pp. 374–380, 2016.

[49] S. Collishaw and et al., "How far are associations between child, family and community factors and child psychopathology informant-specific and informant-general?" *Journal of Child Psychology and Psychiatry*, vol. 50, no. 5, pp. 571–580, 2009.

[50] S. Do-Lenh and et al., "Task performance vs. learning outcomes: a study of a tangible user interface in the classroom," in *European conference on technology enhanced learning*. Springer, 2010, pp. 78–92.

[51] M. K. Crossman and et al., "The influence of a socially assistive robot on mood, anxiety, and arousal in children." *Professional Psychology: Research and Practice*, vol. 49, no. 1, 2018.

[52] C. V. Smedegaard, "Reframing the role of novelty within social hri: from noise to information," in *ACM/IEEE HRI 2019*. IEEE, 2019, pp. 411–420.