JYX JYVÄSKYLÄN YLIOPISTO UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Peltola, Johanna; Kaipainen, Kirsikka; Keinonen, Katariina; Kiuru, Noona; Turunen, Markku

Title: Developing A Conversational Interface for an ACT-based Online Program : Understanding Adolescents' Expectations of Conversational Style

Year: 2023

Version: Published version

Copyright: © 2023 Copyright held by the owner/author(s).

Rights: CC BY 4.0

Rights url: https://creativecommons.org/licenses/by/4.0/

Please cite the original version:

Peltola, J., Kaipainen, K., Keinonen, K., Kiuru, N., & Turunen, M. (2023). Developing A Conversational Interface for an ACT-based Online Program : Understanding Adolescents' Expectations of Conversational Style. In M. Lee, C. Munteanu, M. Porcheron, J. Trippas, & S. T. Völkel (Eds.), CUI '23 : Proceedings of the 5th International Conference on Conversational User Interfaces (Article 1). ACM. https://doi.org/10.1145/3571884.3597142

Developing A Conversational Interface for an ACT-based Online Program: Understanding Adolescents' Expectations of Conversational Style

Johanna Peltola*
Faculty of Information Technology and Communication Sciences, Tampere University
johanna.peltola@tuni.fi

Kirsikka Kaipainen
Faculty of Information Technology and Communication Sciences, Tampere University
kirsikka.kaipainen@tuni.fi

Katariina Keinonen
Department of Psychology, University of Jyväskylä
katariina.keinonen@jyu.fi

Noona Kiuru
Department of Psychology, University of Jyväskylä
noona.h.kiuru@jyu.fi

Markku Turunen
Faculty of Information Technology and Communication Sciences, Tampere University
markku.turunen@tuni.fi

ABSTRACT

A preventative approach is crucial for adolescents' mental well-being, as problems often arise at a young age. Acceptance and Commitment Therapy (ACT) is an evidence-based intervention approach used to enhance psychological flexibility, a central factor in adolescents' mental well-being. Conversational interfaces are recently being experimented with in mental health promotion. Their conversational style plays a significant role in creating meaningful experiences to achieve positive intervention outcomes. In this study, our objective was to understand adolescents' expectations of the conversational style of a text-based virtual coach being developed as part of an ACT-based online program to support intervention engagement. We evaluated eight conversation scripts by collecting qualitative and quantitative data through an online survey from over 200 adolescents. Our findings provide insights on preferred conversational interface features regarding conversational style, including language use, artificiality, and empathy in the domain of adolescent mental well-being.

CCS CONCEPTS

• **Human-centered computing**; • **Human-computer interaction (HCI)**; • **HCI design and evaluation methods**; • **User studies**;

KEYWORDS

conversational interface, conversational style, user expectations, mental well-being, psychological flexibility, acceptance and commitment therapy

*Corresponding author.

This work is licensed under a Creative Commons Attribution International 4.0 License.

CUI '23, July 19–21, 2023, Eindhoven, Netherlands
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0014-9/23/07.
https://doi.org/10.1145/3571884.3597142

ACM Reference Format:
Johanna Peltola, Kirsikka Kaipainen, Katariina Keinonen, Noona Kiuru, and Markku Turunen. 2023. Developing A Conversational Interface for an ACT-based Online Program: Understanding Adolescents' Expectations of Conversational Style. In *ACM conference on Conversational User Interfaces (CUI '23), July 19–21, 2023, Eindhoven, Netherlands.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3571884.3597142

1 INTRODUCTION

Adolescence can be a challenging period in life, with substantial developments in personal identity, emotional experiences, and social life [10]. Research has shown that the peak age for the onset of any mental health disorder is 14.5 years, and approximately 14% of adolescents between 10-19 years of age experience a mental health problem [47, 55]. The availability and accessibility of counseling services often do not meet the increasing need: a preventative approach is a vital part of the solution. To complement traditional one-on-one counseling services or to provide self-guided preventative support for mental health, digital technology applications in the domain of adolescent mental health and well-being have become more widespread, including web-based interventions, mobile applications, serious games, and virtual reality applications [20, 32]. Web-based interventions to promote adolescents' mental health show promise, but self-guided interventions with no human support often suffer from low adherence and high dropout [20, 32, 52].

Text-based conversational interfaces, also called chatbots or conversational agents, have a long history in human-computer interaction. Conversational interface refers to a text-based or voice-based user interface that allows interaction in natural language based on conversational turns, as in human-human communication [35]. While conversational interfaces and chatbots have been studied as a stand-alone support for adolescents [e.g., 18, 50] with promising adherence rates [52], they have been proposed as one possible solution to also increase user engagement in web-based mental health interventions [43, 44]. However, to our knowledge, the role of conversational style used in conversational interfaces to promote intervention engagement has been less studied in the context of

adolescent mental health promotion. A preventative approach also appears less studied in the research about conversational agents in health context [7].

In this paper, we report a study conducted during the development phase of a conversational interface, i.e. "a virtual coach", as a component for a guided online psychological intervention. Youth Compass is a structured and interactive web-based intervention program developed to support adolescent psychological flexibility and well-being by taking the participants through exercises and activities over five weeks. The content is based on Acceptance and Commitment Therapy (ACT), an evidence-based psychological approach stemming from Cognitive Behavioral Therapy (CBT), where acceptance and mindfulness-based techniques are used together with commitment and behavior change strategies [25, 26]. The aim of the program is to promote psychological flexibility by teaching life skills to young people and helping them deal with various challenges and demands of everyday life. Psychological flexibility refers to the ability to commit to values-based actions and to accept inner experiences, that research has shown to be crucial factors in preventing and alleviating mental health issues and supporting psychological well-being among children and adolescents [15].

The present study addresses the following research question: *"How do adolescents evaluate the conversational style related features of a conversational interface aimed at supporting intervention engagement?"* In addressing this question, we provide a deepened understanding of adolescents' expectations of conversational interfaces in mental health promotion and prevention. The findings can benefit both designers and researchers who wish to enhance the quality of human-computer interaction in the context of promoting adolescent mental well-being.

2 RELATED WORK

In this section, we present an overview of research on conversational interfaces for mental health and well-being, including ACT-based approaches, and cover related work about conversational style in conversational interfaces.

2.1 Conversational interfaces for mental health

Most psychological web-based interventions worldwide, including many of the conversational agents for mental health such as Woebot [57] and Wysa [56], are based on Cognitive Behavioral Therapy (CBT) [22, 27]. Acceptance and Commitment Therapy (ACT) is a "new wave" of CBT and has also been recently used in conversational agents, such as Kai.ai [50]. Some conversational agents use several different treatment modalities.

Research in conversational agents for mental health has shown promising findings, but mechanisms of action are unclear [19] and robust evaluation studies are scarce [7]. A recent review [19] identified 13 studies in which conversational agents had been used in the context of treatment of mental health problems, with most studies reporting reductions in psychological distress. However, the influence of a conversational agent's design on participant engagement or intervention outcomes was not assessed in the reviewed studies, and the therapeutic basis for the agents varied [19]. Studies were also heterogeneous in terms of problems addressed, including depression, anxiety, phobia, loneliness, psychological distress, and generally improving mental well-being. Similarly, two other reviews [1, 49] have noted that while chatbots show potential in the delivery of mental health care, their change-promoting mechanisms are not well understood. Moreover, most of the studies covered by these reviews were conducted with adult populations.

Few studies have focused on examining user experience and engagement with conversational agents for mental health. Dosovitsky et al. analyzed interactions with a text-based mental health chatbot called Tess that is composed of 12 modules that utilize multiple treatment modalities such as CBT, ACT and psychoeducation to support users emotionally [12]. Tess was perceived as useful and supportive in a feasibility study conducted with 23 adolescents [48]. In the study, the analysis of module use showed that the ACT-based values module had the highest time spent by the users, although the module was used only by 11% of the users [12]. However, the study could not infer reasons for utilization of different modules of Tess, and the demographics of participants were not collected. User experience data collected from 13 adolescents in a one-session study of Beth Bot, a CBT-based chatbot for depression, suggests that some adolescents find mental health chatbots acceptable but that chatbots should be able to give personalized responses [11]. Kai.ai is a chatbot that has been designed with the broader aim of mental health promotion, utilizing ACT as well as adapting tools from positive psychology [38, 50]. A longitudinal study over a period of four months indicated an increase in the well-being of participants, but the study did not provide insights into how or why the increase took place, and the disengagement from the use of Kai.ai was high [50]. Thus, more in-depth research in user experiences and expectations of conversational interfaces in the domain of mental well-being is warranted.

In the context of promotion of mental health and well-being, only some studies have investigated the experiences of emotional support from a chatbot among adolescents. A two-week study of the CBT-based Woebot used by 16 participants (16-21 years old) focused specifically on how young people experienced different types of social support [4], categorized into appraisal, emotional, informational and instrumental support according to House et al. [29]. In ACT context, a particularly relevant form of social support is appraisal support that manifests e.g., as feedback, social comparisons and affirmations that are relevant for a person's self-evaluation. The findings of the Woebot study suggest that emotional support requires designing a conversational interface to communicate in a more humanlike manner than informational or appraisal support [4].

KIT is a text-based chatbot that has been designed to support positive body image among people with eating issues, and its content is based on psychoeducation, CBT, ACT, and mindfulness [5]. In a qualitative online focus group study, it was found that adolescents appreciated KIT's non-gendered, non-human and cheerful appearance, but criticized its lengthy content chunks and formal language style [5]. The study also indicated that inserting pictures of the chatbot character into the conversation made participants feel they were having a conversation with a real 'someone', even though they were aware that the 'someone' was a programmed entity.

2.2 Conversational style in conversational interfaces

An experimental study by Mariamo et al. [34] addressed specifically the communication style of a mental health chatbot for adolescents: the findings were inconclusive regarding the tone of voice, with some participants appreciating the friendly tone that made the chatbot more relatable, but some criticizing an 'overly friendly' chatbot as 'trying too hard' [34]. Conferring to the findings of the KIT chatbot study [5], adolescents appear to prefer a conversational interface with an informal and friendly tone. In another study about a mental health chatbot for young people, it was found that adolescents wanted the chatbot to use adolescent slang [23]. Fadhil et al. discovered that using emojis instead of plain text made the study participants, mainly young adults, more confident in sharing information about their mental well-being, but the result was inverted when discussing physical well-being [14]. The findings regarding the use of emojis by chatbots for adolescents are mixed, as emojis were preferred by adolescents in a study about chatbot co-development [23], but negatively perceived in a more recent study about chatbot user testing [11].

In the domain of adolescent sexual and reproductive health, Rahman et al. [40] evaluated a chatbot prototype with adolescents, university first-year students and medical personnel. In terms of communication style, adolescents seemed to appreciate the chatbot providing authentic information, asking counter-questions from them, and giving concise answers. Pragmatic attributes of conversational agents, such as how useful the content is perceived by the user, are important factors in conversation design [17]. However, the quality of information is not everything: hedonic aspects of experience, such as fun and pleasure, are essential to consider when designing for user engagement [51]. Interestingly, younger people highlighted hedonic aspects more often than older participants in a questionnaire study of users' experiences with chatbots by Følstad & Brandtzaeg [17].

It is well established that people tend to apply humane characteristics and social rules to computers as with other people in real-life social situations, known as the CASA (Computers as social actors) paradigm [39, 41]. More humanlike is not necessarily better when it comes to conversational interfaces: in a study regarding emotional needs of teenagers, the agent's lack of emotion was perceived as beneficial for being a good, non-judgmental listener the user could confide in [30]. A study on Vivibot, a chatbot for promoting psychological well-being after cancer treatment, had similar results among young people in terms of having the bot as a non-judgmental listener [21]. In a study by Lucas et al. [33] concerning health screenings, adult participants were more willing to self-disclose to a virtual interlocutor: their fear of being evaluated was lower than with a human interlocutor. Humanlike features may also backfire due to a phenomenon known as the uncanny valley where an artificial agent causes an unpleasant emotional reaction due to being very humanlike, but not quite there [36]. However, Skjuve et al. [45] found no support that text-based conversational agents would likely cause an uncanny valley effect. Moreover, prior research indicates that conversational agents that disclose information about themselves are better received by users, increasing engagement and self-disclosure [6, 31, 33]. In a study by Bickmore et al. [6], participants did not perceive a text-based agent as dishonest when it disclosed personal information, even though it could not be true for a virtual entity.

Research suggests that empathy and relational behavior displayed by a conversational agent strongly contributes to positive user experience in virtual health assistants, helping to build trust and rapport [9]. Perception of empathy also contributed to positive user experience among young adults in a study concerning Woebot, a text-based conversational agent delivering CBT [16]. However, displaying empathy in a text-based interface is more difficult without an avatar displaying nonverbal, empathetic behaviors, and must rely only on verbal content [9]. In the present study, we investigate how adolescents perceive the conversational style of a text-based virtual coach in terms of empathy, artificiality, and friendliness, among other features.

3 METHODS

We conducted a mixed-method questionnaire study in the context of an online event aimed at adolescents. We selected eight pre-generated conversation scripts between a virtual coach and an interlocutor from the Youth Compass program in development for evaluation. In this section, we describe the participants, study procedure, conversation scripts, survey questions and data analysis.

3.1 Participants

The main target group of the program is Finnish-speaking ninth-graders. To reach our target population, we collected the data during an online event targeted at students attending the ninth grade of Finnish secondary schools in Tampere, Finland. Permission to conduct research on adolescent participants was granted by the city of Tampere. The participants' approximate age was 15 years, and they attended four different secondary schools. Age was not collected in the survey as the approximate age of the students was known.

Circa 216 students participated in the study. Of all 419 responses, 50.6% were from female participants, and 38.4 % from male participants. The rest were from participants identified as "other" (2.1%) and those who did not share their gender (8.8%). The number of individual respondents is approximate due to the nature of the data collection: two optional surveys were available for the participants to complete each day over the course of the four-day event. A participant could choose to complete one of the surveys and skip the other, or to complete both surveys.

3.2 Procedure

Vaikuta! is an online event that roughly translates to *Make an impact!* from Finnish. The purpose of the event that was repeated on four separate days was to allow the participants to familiarize themselves with societal activities and to voice their opinions on real-life issues. The event was arranged on a digital education platform Seppo.io where the students participated by playing a game that consisted of completing surveys, creative assignments, and sharing opinions. After each activity, the participants were rewarded with points counting towards a final score in the game. The participants had 1.5 hours to play the game, during which they

Table 1: Logged-in players, completed surveys, and the response rates.

Day	Logged-in players	Responses to survey 1	Response rate	Responses to survey 2	Response rate
1	14	7	50%	7	50%
2	101	62	61%	57	56%
3	70	53	76%	52	74%
4	117	83	71%	80	68%
	302	**205**		**196**	

Table 2: Youth Compass program module themes, conversation script (CS) topics, their descriptions, and the day of the event when they were assigned to study participants.

Module #	Module theme	CS #	CS Topic	CS Description	Event Day
1	Values and values-based actions	1A	Goal setting	Identifying the user's values-based goals and planning values-based action.	2
1	〃	1B	Life satisfaction	Recognizing what is valuable in daily life and exploring if changes are warranted.	4
2	Acceptance	2A	Unpleasant thoughts	Acknowledging that experiencing unpleasant thoughts is a common experience.	3
3	Defusion: observing and accepting thoughts	3A	Social media and comparison	Social media use and questioning self-critical thoughts that may arise from comparing oneself to others.	1
3	〃	3B	Social media and inspiration	Recognizing which social media content evokes emotional responses, and if the user wishes to seek those things in their personal life.	4
4	Self-compassion	4A	Self-talk	Exploring and questioning self-critical thoughts and attitudes.	3
5	Relationships and prosocial behavior	5A	Academic performance	Noticing perfectionistic strivings and stress and anxiety related to schoolwork.	2
5	〃	5B	Self-compassion	Contrasting compassion towards others and self-compassion.	1

could freely choose the activities they wanted to complete. They were not required to complete all the activities.

According to the event organizer's guidelines, one activity was allowed to take approximately 5 minutes. We designed two surveys for each event day, both including one conversation script, altogether eight surveys. They were presented on the Seppo.io platform as separate activities: each participant could choose to complete one, both, or neither of the two surveys. Each day was dedicated to one of the four student groups attending different schools. The average response rates for the first and the second survey were 64.5% and 62%, respectively (see Table 1).

On the first day of the event, only one class of students participated in the game instead of four classes planned beforehand. This resulted in a significantly smaller number of survey respondents on the first day affecting two surveys that were assigned on that day (see Table 2).

Before the event commenced, the first survey was completed by 11 respondents, and the second survey by 7 respondents for testing purposes to ensure that the questions were understandable, pictures appeared correctly, and that the online form was working properly. The surveys that were completed before the actual event were counted in the total number of 425 completed surveys. Six were removed as they were deemed to be duplicates, resulting in 419 responses.

3.3 Virtual coach and the conversation scripts

The virtual coach (VC) of the Youth Compass program is a chatbot character named Rami, with a static fox avatar. We selected a non-human avatar as a neutrally aligned character was preferred in testing with a pilot group of four adolescents. We assumed that an animal, such as a fox, would be more neutral than a human

character, that could potentially alienate some users due to identifiable human-like features [cf. 5]. The VC is designed as a rule-based conversational interface that allows the user to respond through predefined responses. It also allows a limited amount of unconstrained input at certain stages of the program. During the data collection, the VC was being developed and the current scripts do not represent the final conversation scripts implemented in the online program as the findings presented in the current paper were used to iterate the conversation designs before their technical implementation.

The purpose of the VC is to guide the users through the exercises and encourage them to complete the program while providing appraisal support [cf. 4]. The conversational style of the scripts was designed to reflect the content and goals of exercises and other content already implemented in the online program. The script drafts used in the current analyses were developed in an iterative manner and at the time of the event we were focused on the conversational style of the VC. For the purposes of this study, we selected eight hypothetical, noninteractive conversation scripts (presented in Figures 1 to 8 translated from the original language) between the VC and an interlocutor. There was no actual interaction between the VC and an interlocutor as the scripts were pre-generated. The content of the scripts, i.e. the VC responses and the predefined responses for the interlocutor, was designed by a group of psychologists at University of Jyväskylä based on ACT. We selected the script topics to adequately represent the themes of the Youth Compass program in our study. The content of the scripts follows the content of the five modules in the program (see Table 2). The themes of the modules are typical in ACT-based programs and approaches, which promote values clarification, value-based actions, mindfulness, acceptance, and defusion to increase psychological flexibility [15]. Likewise, the individual scripts follow the typical content of an individual exercise or task in an online ACT-program. The content reflects themes that were previously presented in a traditional form in the online program, e.g., in the form of a written exercise, an audio exercise or a behavioral task.

3.3.1 Design rationale of the virtual coach's conversational style. In all scripts, the VC's conversational style was designed to feel as natural as possible, as if talking with a real coach. The language was aimed to be kept simple and clear to avoid confusion, and to keep an informal and colloquial tone for easier relatability. The coach sometimes uses emojis to enhance the message. Throughout all scripts, the coach was always designed to seem friendly, empathetic and non-judgmental toward the interlocutor. To keep the conversation interesting, the coach sometimes discloses their own personal experiences and asks the interlocutor specific questions about their life related to the topic at hand, to make the conversational exchange more personal and allow the interlocutor's self-reflection. The scripts were all 1-2 turns long to keep them concise for the survey. Below, we provide further details of individual conversation scripts and the design rationale behind them.

Goal setting (1A). In 1A, VC is designed as encouraging and enthusiastic to support the interlocutor in setting goals and making them excited about it.

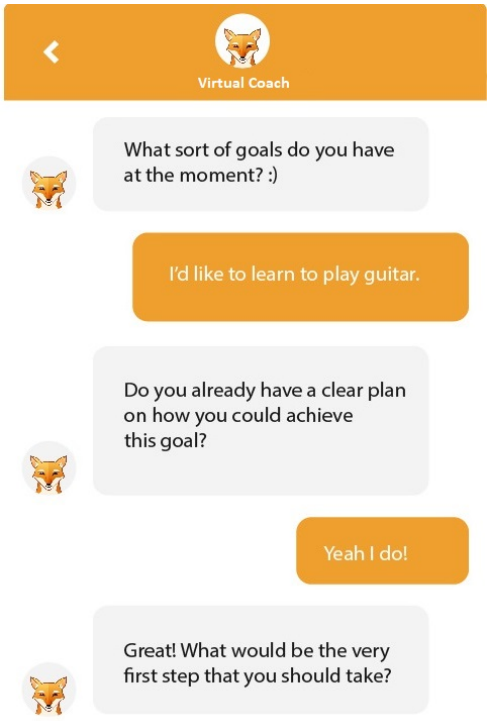

Figure 1: Goal setting script (1A).

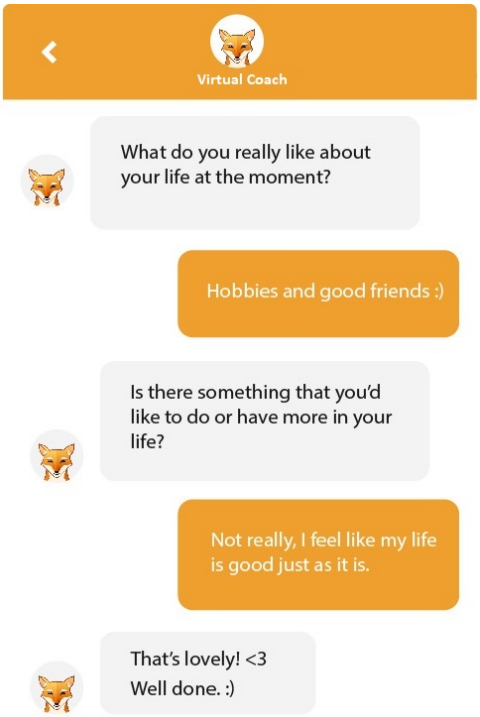

Figure 2: Life satisfaction script (1B).

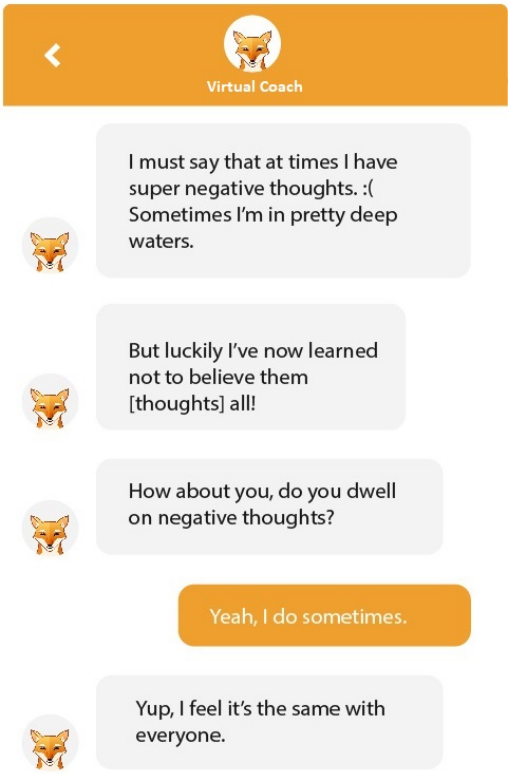

Figure 3: Unpleasant thoughts script (2A).

Life satisfaction (1B). In 1B, the goal of the conversational style is similar as in 1A, and the VC is aiming to guide the interlocutor in exploring their satisfaction with current life situation and to provide encouraging praise when the interlocutor says they are happy with their life.

Unpleasant thoughts (2A). In 2A, the VC aims to provide peer support to the interlocutor while sharing its own experience of having unpleasant thoughts. Hence, the VC's conversational style is designed as slightly moody for the interlocutor to relate to, and perhaps encourage them to open about their unpleasant thoughts too. By providing a comment about everyone having similar thoughts, the VC aims to bring up an important point of how common it is to have such thoughts, possibly providing relief to the interlocutor.

Social media and comparison (3A). In 3A, the VC's conversational style is designed as reflective, guiding the interlocutor to consider their social media use and its influence on them.

Social media and inspiration (3B). In 3B, the VC adopts a reflective and even slightly inquisitive conversational style as it tries to elicit information from the interlocutor.

Self-talk (4A). In 4A, the VC is designed as enthusiastic and giving peer support to the interlocutor. With its comment about not having to believe one's own thoughts, it aims to introduce the concept of defusion to the interlocutor in a supportive manner.

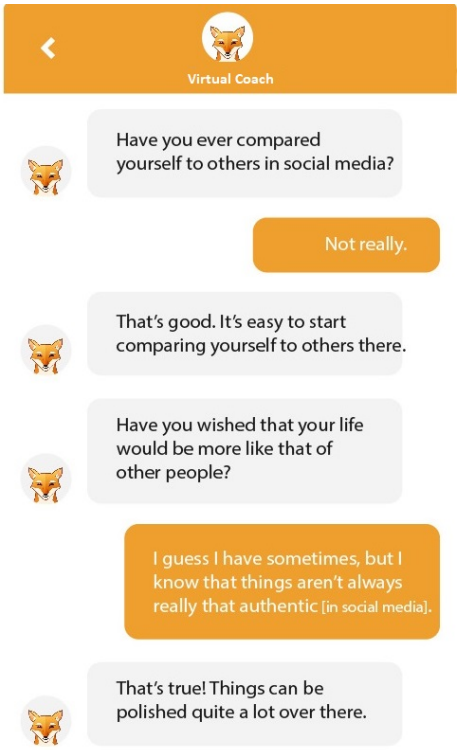

Figure 4: Social media and comparison script (3A).

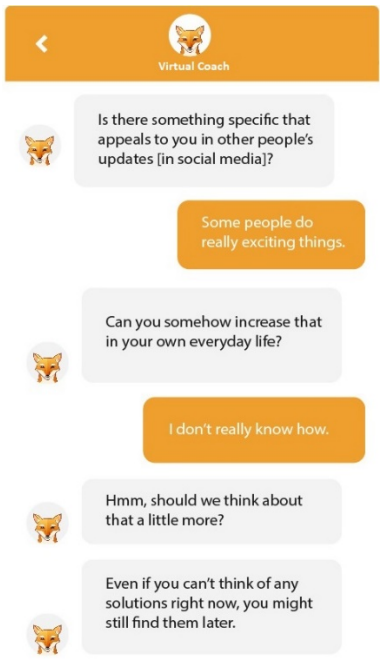

Figure 5: Social media and inspiration script (3B).

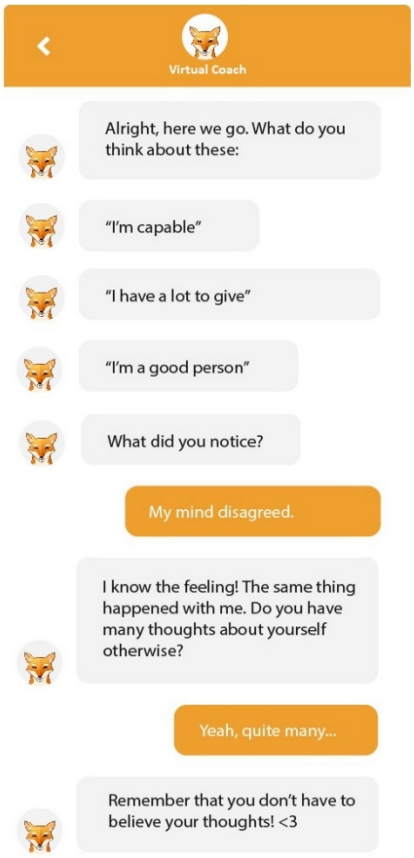

Figure 6: Self-talk script (4A).

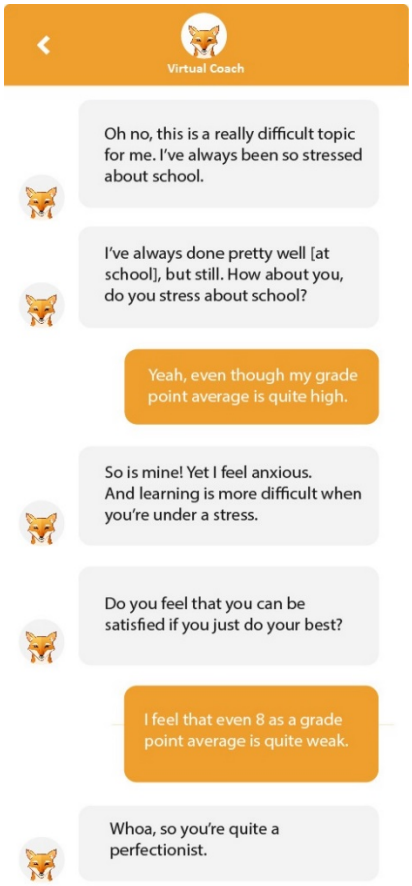

Figure 7: Academic performance script (5A).

Academic performance (5A). In 5A, the VC is offering peer support by sharing its own experiences, and the approach is rather distressed as the VC expresses its strong feelings related to the topic of academic performance. This is to allow the interlocutor to relate to the shared experience and encourage them to share their genuine feelings on the topic.

Self-compassion (5B). Here, the VC aims to provide empathy and understanding to the interlocutor regarding the topic of self-compassion, while using an example on the notion that it is often easier to be compassionate toward others than oneself.

3.4 Survey questions

Each survey contained an image of one conversation script that was evaluated by the participants. Conversation scripts are described and displayed in section 3.3.

To evaluate participants' expectations regarding the conversational style of the virtual coach, we collected quantitative data with semantic differential and qualitative data with an open-ended question. Semantic differential is an established and well-validated method for measuring the respondent's perceptions of things and concepts [46]. Respondents can indicate their experience of a product or a service on a five or seven-point scale that consists of bipolar evaluative dimensions. The participants were guided to read the script and then evaluate the VC's utterances on a five-point semantic differential scale across six dimensions: *Interesting – Boring, Empathetic – Indifferent, Clear – Confusing, Encouraging – Discouraging, Friendly – Rude,* and *Natural – Artificial.* The survey items were selected to reflect the desired characteristics of the VC and its conversational content. A higher numerical value was assigned to the positive endpoint of the scale, e.g., *5=Clear* and *1=Confusing*. Cronbach's Alpha value for the six survey items was $\alpha = .882$.

The participants were also asked to elaborate on their impressions of the virtual coach in an open-ended question with the following prompt: *"Please elaborate how Rami sounds to you and what do you think about Rami's utterances."* All survey content, responses and participant quotations have been translated from Finnish to English.

3.5 Data analysis

Responses to the semantic differential were entered into IBM SPSS Statistics (Version 28.0.1.0) [53] for descriptive analysis. The first author also conducted a manual sentiment analysis to the open-ended qualitative evaluations, applying a label of positive, neutral or negative to each response containing a valid input. The labels and responses were then collated as a list and discussed between three

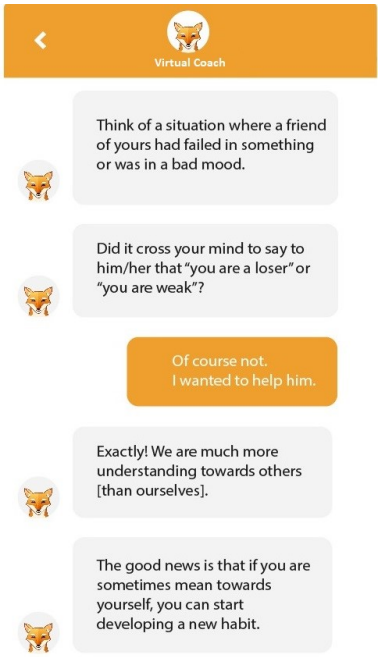

Figure 8: Self-compassion script (5B).

researchers to reach a common understanding of the expressed sentiments, after which it was decided which label to use for any given response. A positive label was applied if the participant's response consisted of mostly positive reactions. A neutral label was applied if the reactions were ambiguous or contained only neutrally aligned reactions. Finally, a negative label was applied if the response contained mainly negative reactions. Finally, the first author conducted a qualitative content analysis of the open-ended responses to reflect the participants' perceptions of the virtual coach's conversational style. To get acquainted with the data, the first author repeatedly read all the responses to the open-ended question and then created the initial codes and categories using inductive coding [13]. As the content analysis progressed, the codes and categories were revised as needed. To increase the validity of the analysis, the participant quotations and the codes and categories assigned to them were collated as a list and sent to two other authors for cross verification. The codes and categories were discussed between three researchers in various stages of the analysis to find a common understanding and select the final interpretation. ATLAS.ti Windows (Version 22.1.5.0) [54] software was used to assist in coding and interpreting the data.

4 RESULTS

In this section, we present the results for the quantitative and qualitative conversation script evaluations. A total of N=419 survey responses were analyzed. We present the results of the semantic differential, sentiment analysis, and the content analysis.

4.1 Quantitative evaluations of conversation scripts

In the overall semantic differential results (Figure 9), friendliness and artificiality of the VC stood out. Across all conversation scripts, the *Natural – Artificial* dimension received the lowest mean rating of 3.09 (SD=1.14) whereas the *Friendly – Rude* dimension received the highest rating of 3.89 (SD=1.06).

Self-compassion script (5B) was evaluated as the most interesting, encouraging and natural. ***Social media and comparison*** (3A) was evaluated as the most friendly and clear. ***Academic performance*** (5A) was rated as the least friendly. Unlike in most of the other scripts, female participants found the virtual coach less encouraging than male participants in ***Academic performance*** (5A) and ***Unpleasant thoughts*** (2A). The coach was considered the most empathetic in ***Goal setting*** (1A). ***Unpleasant thoughts*** (2A) was rated the least interesting, empathetic, encouraging and natural.

The highest mean rating for both male and female participants was in the *Friendly – Rude* dimension (3.63 and 4.07 respectively), and the lowest rating in the *Natural – Artificial* dimension (2.82 and 3.26 respectively). All dimensions were rated higher by female participants.

4.2 Sentiment analysis of qualitative evaluations

All participant responses to the open-ended question *"Please elaborate how Rami sounds to you and what do you think about Rami's utterances."* were analyzed in a manual sentiment analysis where a label of positive, neutral or negative was applied to each response (see Table 3).

Sentiment analysis revealed differences in participants' reactions between various scripts. Also gender differences were evident. Female participants had more positive sentiments (45.8%) in their qualitative evaluations than male participants (34.8%). Male participants also had more negative sentiments (37.3%) than female participants (30.2%).

Conversation scripts ***Goal setting*** (1A), ***Self-talk*** (4A) and ***Life satisfaction*** (1B) divided opinions among the participants: female participants reacted more positively to them whereas ***Social media and inspiration*** (3B) was received better by the male participants. ***Goal setting*** (1A) had the most positive overall response. In response to ***Academic performance*** (5A), female participants gave more negative sentiments (43.5%) than male participants (27.3%). ***Unpleasant thoughts*** (2A) received the highest number of negative sentiments (47.2%) as well as the least number of positive sentiments (18.9%) among all scripts. See Table 4 for all the results.

4.3 Content analysis of qualitative evaluations

4.3.1 Most common impressions. The participants' qualitative evaluations of the virtual coach were coded for further analysis of the data. The most often found positive impression (N=71) was related to the coach being **friendly**, nice or approachable, which were all placed under the same coding category. The most often found negative impression (N=69) was related to the coach being **artificial**, unnatural or robotic. We could also see other concurrences with the semantic differential results, such as natural and empathic (see Table 5).

Table 3: Sentiment analysis labels, their descriptions, example quotations and frequencies.

Label	Description	Example quotation	Frequency
Positive	A mostly positive perception of the VC	"Rami sounds friendly and encouraging." (5B)	168 (44.1%)
Neutral	A neutral or ambiguous perception of the VC	"[The VC] tells about things and the words are quite okay." (2A)	80 (21%)
Negative	A mostly negative perception of the VC	"Rami's words seem rude and indifferent."	133 (34.9%)

Table 4: Sentiment analysis results for each conversation script.

Conversation script	Sentiment: pos N (%)	Sentiment: neutral N (%)	Sentiment: neg N (%)	Missing sentiments
1A: Goal setting (N=57)	35 (61.4%)	4 (7%)	13 (22.8%)	5 (8.8%)
Female (N=25)	18 (72%)	2 (8%)	4 (16%)	1 (4%)
Male (N=29)	15 (51.7%)	2 (6.9%)	9 (31%)	3 (10.3%)
Other (N=2)	1 (50%)	0	0	1 (50%)
No answer (N=1)	1 (100%)	0	0	0
1B: Life satisfaction (N=83)	32 (38.6%)	19 (22.9%)	26 (31.3%)	6 (7.2%)
Female (N=49)	23 (47%)	11 (22.4%)	12 (24.5%)	3 (6.1%)
Male (N=27)	5 (18.5%)	7 (25.9%)	13 (48.1%)	2 (7.4%)
Other (N=2)	1 (50%)	1 (50%)	0	0
No answer (N=5)	3 (60%)	0	1 (20%)	1 (20%)
2A: Unpleasant thoughts (N=53)	10 (18.9%)	17 (32.1%)	25 (47.2%)	1 (1.9%)
Female (N=30)	5 (16.7%)	10 (33.3%)	14 (46.7%)	1 (3.3%)
Male (N=21)	5 (23.8%)	6 (28.6%)	10 (47.6%)	0
No answer (N=2)	0	1 (50%)	1 (50%)	0
3A: Social media and comparison (N=18)	5 (27.8%)	10 (55.6%)	2 (11.1%)	1 (5.6%)
Female (N=5)	3 (60%)	2 (40%)	0	0
Male (N=1)	0	1 (100%)	0	0
Other (N=1)	0	0	1 (100%)	0
No answer (N=11)	2 (18.2%)	7 (63.6%)	1 (9.1%)	1 (9.1%)
3B: Social media and inspiration (N=80)	32 (40%)	14 (17.5%)	26 (32.5%)	8 (10%)
Female (N=48)	19 (39.6%)	9 (18.8%)	15 (31.3%)	5 (10.4%)
Male (N=26)	12 (46.2%)	4 (15.4%)	7 (26.9%)	3 (11.5%)
Other (N=1)	0	0	1 (100%)	0
No answer (N=5)	1 (20%)	1 (20%)	3 (60%)	0
4A: Self-talk (N=52)	25 (48.1%)	2 (3.8%)	21 (40.4%)	4 (7.7%)
Female (N=27)	17 (63%)	0	9 (33.3%)	1 (3.7%)
Male (N=23)	7 (30.4%)	2 (8.7%)	11 (47.8%)	3 (13%)
No answer (N=2)	1 (50%)	0	1 (50%)	0
5A: Academic performance (N=62)	24 (38.7%)	14 (22.6%)	19 (30.6%)	5 (8.1%)
Female (N=23)	8 (34.8%)	4 (17.4%)	10 (43.5%)	1 (4.3%)
Male (N=33)	12 (36.4%)	9 (27.3%)	9 (27.3%)	3 (9.1%)
Other (N=2)	1 (50%)	1 (50%)	0	0
No answer (N=4)	3 (75%)	0	0	1 (25%)
5B: Self-compassion (N=14)	7 (50%)	2 (14.3%)	1 (7.1%)	4 (28.6%)
Female (N=5)	4 (80%)	0	0	1 (20%)
Male (N=1)	0	0	1 (100%)	0
Other (N=1)	1 (100%)	0	0	0
No answer (N=7)	2 (28.6%)	2 (28.6%)	0	3 (42.9%)

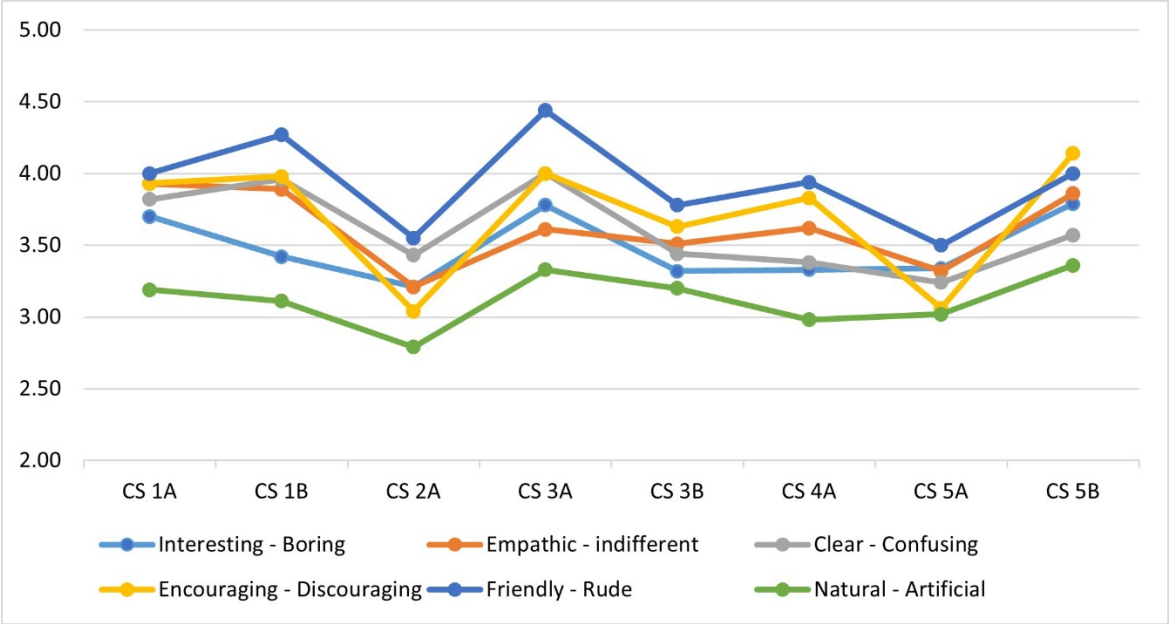

Figure 9: Semantic differential results for each conversation script (CS).

Table 5: Overview of the most common impressions of the virtual coach in the qualitative evaluation.

Impression	Example quotation	Frequency
Friendly	"He sounds nice and friendly." (5A)	71
Artificial	"Sounds like a machine." (2A)	69
Good wording	"Rami's words were really good." (3B)	57
Encouraging	"It's encouraging, but not too enthusiastic." (1A)	44
Natural	"Rami sounds as if you're talking with your own friend." (1A)	42
Empathic/interested	"Rami sounded very empathic." (3A)	41
Unempathetic/self-centered	"Rami sounds like they don't care about other people's affairs, but their own." (2A)	29
Unhelpful	"Rami's diction is transient and not helpful in any way." (2A)	25
Forced	"I think Rami's words are not appropriate, but they seem forced somehow." (3B)	19

4.3.2 Language use. Table 6 presents the results of the content analysis related to the evaluations of language use in the scripts, with example quotations. Several participants (N=14) preferred the virtual coach to talk in colloquial language, which was mentioned most about ***Life satisfaction*** script (1B), whereas only two participants preferred standard language. Some participants (N=9) said that the VC is trying to imitate the way young people talk which contributed to the perception of artificiality. Remarks related to grammar and vocabulary were mostly given by female participants (N=11) across all scripts, while only one male participant made such a remark. Emojis evoked positive feelings in some and negative in a few participants. Too many or too long messages were criticized by some participants, especially in ***Unpleasant thoughts*** (2A) where the coach shares its experiences of negative thinking. In ***Life satisfaction*** (1B), six participants (N=6) mentioned that the VC should have further elaborated its questions by giving more details or asking further questions or giving further comments regarding the subject of conversation.

4.3.3 Conversational style. Table 7 presents conversational style results with example quotations from participants. ***Goal setting*** (1A) had the most mentions of the coach being empathetic and caring (N=14), especially by female participants (40%). Of all participants evaluating this script, 19.3% said that the coach seemed interested in the conversation. However, four participants (7%) found the VC too inquisitive with its questions. The coach's encouraging approach divided opinions in 1A and was also considered artificial (N=6) and even arrogant by one respondent.

In ***Life satisfaction*** (1B), six participants found the coach to resemble an authority figure such as a teacher or a school psychologist, and five viewed the VC as boring, more than in any other script. Another six participants found the coach's comment *"That's lovely!*

Table 6: Content analysis results in language use.

Theme	Example quotation	Frequency	Target scripts
Colloquial language preferred	"Modern vernacular, which is good! Sounds like a lovely normal young person." (2A)	N=14	1A (N=1), 1B (N=8, 9.6%), 2A (N=2), 3A (N=1), 3B (N=1), 5A (N=1)
Standard language preferred	"I think Rami could have talked a bit differently. If Rami is trying to help, then standard language wouldn't hurt in that situation. Some words just didn't fit in." (3A)	N=2	1B (N=1), 3A (N=1)
Emoji preferred	"I think incorporating emojis such as "<3" and ":)" into the text emphasizes friendliness and empathy, so it doesn't seem so rude." (1B)	N=4	1A (N=1), 1B (N=2), 4A (N=1)
Emoji not preferred	"A heart from a bot that cannot feel love is a bit too much." (1B)	N=2	1A (N=1), 1B (N=1)
Grammar or vocabulary not liked	"Rami's word choices are quite declarative and generic. Rami sounds friendly, but in the end quite neutral ('that's good', 'quite a lot'). I think Rami could be even more encouraging with its words." (3A)	N=14	1A (N=1), 1B (N=2), 2A (N=3), 3A (N=3, 16.7%), 3B (N=2), 4A (N=1), 5A (N=2)
Correct grammar makes the VC seem artificial or official	"Too correct grammar in the messages makes the person sound a little too official." (2A)	N=2	3A (N=1), 4A (N=1)
Messages are too long or too many	"I don't like it how the messages are too long and too many in a row." (2A)	N=4	2A (N=3), 3B (N=1)
Messages are too short	"Rami sounds rude and indifferent because his responses are short." (5A)	N=2	3B (N=1), 5A (N=1)
The VC should ask more subsequent questions for elaboration	"Rami sounds good, but I think he should ask about things more." (1B)	N=10	1B (N=6, 7.2%), 2A (N=1), 4A (N=1), 3B (N=1), 5A (N=1)
The VC is trying to talk like young people	"Sounds exactly like an adult who tries to talk in youth language, even though there is no such thing these days." (2A)	N=9	1A (N=1), 1B (N=1), 2A (N=3, 5.7%), 3A (N=1), 3B (N=2), 4° (N=1)

<3 Well done. :)" arrogant, too friendly, tacky, or artificial. For one participant it seemed odd that the coach was asking what more would the interlocutor want, when the request was to think about positive things in life. Another participant also said that the coach is trying hard to find something negative from the interlocutor's life.

In ***Social media and inspiration*** (3B), 10 participants (12.5%) described the coach as forced, inquisitive or aggressive, which was the highest among all scripts. One respondent felt that the coach is imposing his agenda on the interlocutor. Also in 3B, 13 respondents (16.3%) found the coach unhelpful, feeling that he was uncapable of providing solutions and even negatively influencing the interlocutor to compare themselves with others in social media. However, five participants found it helpful when the coach asked questions on how to discover and do inspiring things. In ***Life satisfaction*** (1B), ***Unpleasant thoughts*** (2A) and ***Self-talk*** (4A) the coach was considered unnatural or artificial by over 20% of all respondents. Utterances that seemed poorly worded likely contributed to this effect. In 4A, five participants found the coach's comment about not having to believe one's own thoughts unhelpful. Similarly, in 2A, three participants said the coach's comment *"Yup, I feel it's the same with everyone"* was unhelpful. Three participants referred to the coach's personal experiences as disingenuous. One participant liked the coach's messaging style in 4A labeling it "robotic": *"This was somehow really sweet in all its roboticness! Probably no real human would send individual statements to another person and ask to think about them, but it suits Rami the bot very well."*

In ***Academic Performance*** (5A), the VC shared their academic grievances, which many participants disliked: 39.1% of female participants labeled the coach as self-centered, unempathetic or indifferent whereas only 3% of male participants did so. It seemed to trigger impressions of the coach turning too much attention on themself, not being interested in the interlocutor, and even being arrogant or annoying with some of its comments. However, six female participants liked the coach's approach in 5A and two said that the coach was successful in giving peer support to the interlocutor. In ***Unpleasant Thoughts*** (2A) where the coach talks about dealing with negative thoughts, 16.7% of female participants considered the coach as self-centered whereas only 4.8% of male participants did so. Two people specified that the comment about everyone having similar problems made the coach seem unempathetic and indifferent towards the interlocutor.

Table 7: Content analysis results in conversational style.

Theme	Example quotation	Frequency	Target scripts
The VC is empathetic, caring and interested in the interlocutor's affairs	"Rami is clearly interested. With subsequent questions they show their interest and that they have read the messages carefully." (1A)	N=41	1A (N=14, 25.6%), 1B (N=10), 2A (N=1), 3A (N=1), 3B (N=5), 4A (N=4), 5A (N=6)
The VC is self-centered, unempathetic or indifferent	"Rami shouldn't have praised themself when they didn't yet know how the other person fares at school. They shouldn't have said the last comment; I've been told the same and it's really annoying." (5A)	N=29	1A (N=1), 1B (N=3), 2A (N=7), 3A (N=1), 3B (N=2), 4A (N=4), 5A (N=11, 17.7%), 5B (N=1)
The VC is encouraging	"I think Rami is friendly and encouraging, and he tries to help."	N=44	1A (N=10, 17.5%), 1B (N=6), 2A (N=1), 3B (N=12), 4A (N=8), 5A (N=3), 5B (N=4, 28.6%)
The VC is too inquisitive, aggressive or forced	"I think Rami sounded like they want to incorporate these features into the interlocutor's life by force." (3B)	N=19	1A (N=4), 1B (N=1), 3B (N=10, 12.5%), 4A (N=1), 5A (N=3)
The VC is friendly, nice or approachable	"Rami sounded friendly and encouraging." (1B)	N=71	1A (N=13), 1B (N=24, 28.9%), 2A (N=4), 3A (N=5, 27.8%), 3B (N=8), 4A (N=6), 5A (N=9), 5B (N=2)
The VC is helpful	"[Rami sounds] friendly and very helpful when they start to think together with the interlocutor what to do." (3B)	N=9	1A (N=2), 1B (N=1), 3B (N=5, 6.3%), 5A (N=1)
The VC is unhelpful	"It's nice how it gets the young person to ponder, but it also clearly made the person to compare themself to others in social media, which is probably not that good." (3B)	N=25	1A (N=1), 2A (N=3), 3B (N=13, 16.3%), 4A (N=5), 5A (N=3)
The VC is natural or unrobotic	"Sounds very natural. You wouldn't even believe that the respondent is not a real human." (1B)	N=42	1A (N=7), 1B (N=11, 13.3%), 2A (N=4), 3A (N=2), 3B (N=9), 4A (N=3), 5A (N=4), 5B N=2 (14.3%)
The VC is artificial, unnatural or robotic	"Sounds a bit like a robot because some of the responses were quite weird and didn't fit to other person's comments. I think Rami could have elaborated." (4A)	N=69	1A (N=6), 1B (N=17), 2A (N=11), 3A (N=4), 3B (N=8), 4A (N=13, 25%), 5A (N=8), 5B (N=2)
The VC's personal experiences are not convincing	"Sounds quite artificial. It's surprising how much [the VC] talks about itself when you would think that it's a machine." (2A)	N=3	2A (N=1), 4A (N=1), 5A (N=1)

5 DISCUSSION

In this section, we discuss the main findings of the study, the identified issues to consider in future research and design of conversational interfaces for supporting adolescents' mental well-being and intervention engagement, and study limitations.

5.1 Main findings

The central aim of our study was to examine *how adolescents evaluate the conversational style related features of a conversational interface aimed at supporting intervention engagement*. Regarding adolescents' experiences of positive characteristics of the investigated conversational interface, we discovered that while friendliness, empathy, and encouraging character are desired features, they can also be considered artificial by some if used in excess. Empathy seemed to be a valued characteristic in our text-based virtual coach, even if only conveyed verbally without an animated avatar. Participants considered the virtual coach empathetic when it showed interest in the interlocutor's affairs by actively commenting and asking questions.

Regarding adolescents' criticism of the virtual coach, one recurring theme was the perceived artificiality: the reasons mentioned included utterances viewed as poorly worded, formal language, flawless grammar, unnatural messaging style (such as sending several messages in a row) and attempting to imitate the way young people talk. The informal and friendly tone of the virtual coach, such as the use of colloquial language, was preferred by the participants in our study, concurring with results from many other chatbot studies [5, 23, 34].

Previous research has suggested that self-disclosing conversational agents are preferred by users [6, 31, 33], but we also saw

in our study that not everyone thinks this way: some participants perceived the virtual coach's personal examples as artificial, as they could not be true. A natural and humanlike virtual coach may not be something to aim at, as one participant mentioned: *"One could perhaps notice that it's a robot, but I don't know if that's necessarily a bad thing."* Users might be more comfortable with and willing to disclose to a virtual agent that does not resemble another human too much, as previous research has suggested [33]. These somewhat mixed findings about artificiality bear similarities to the findings of a scenario study that found that adolescents perceived a social robot self-disclosing humanlike emotions and experiences as artificial and unnecessary, suggesting that a robot should express itself in a way that is appropriate to a non-human [3].

The perceived helpfulness of the content may depend on the language used: choice of words and sentence formulation might not be appropriate for the content that is otherwise considered beneficial, and hence the message might get misinterpreted. In many instances, participants reported that the virtual coach's wording was somehow off but could not always explain what was wrong: again, individual differences in interpreting the conversational style should be considered. The topic of the ***Unpleasant Thoughts*** script (2A) was acceptance and acknowledging that having unpleasant thoughts is a common experience: the wording of the VC might have come across in an undesired way as some participants said the VC dismissed their problems when he stated that everyone has similar issues, perhaps making the statement sound like a cliché. Moreover, previous research has shown that users are significantly more critical of responses framed as coming from a virtual agent rather than a human interlocutor, even when the responses are identical [37].

In ***Social media and inspiration*** script (3B), many participants found the coach unhelpful, elaborating that it tried to make the interlocutor compare themselves to others. However, the VC's purpose was to give another perspective on how to view others' updates, which can often be a reason for social comparison, and turn things around for a more positive outlook. Although our study only focused on conversational style and not on interventional effects, we may argue that despite the negative feedback, a positive outcome may still ensue: after all, the goal of the program that the VC would be a part of is to enhance adolescents' psychological flexibility and experiencing a variety of emotions is part of the process. Similarly, in ***Unpleasant Thoughts*** (2A), the topic evoked responses that perhaps reflected the participants' own frustration, which does not indicate that the content per se would be flawed. We can surmise that the script content in 2A was not an ideal way to introduce the complex topic of accepting one's unpleasant thoughts, but it is possible that the dialogue would be perceived differently in the context of the program, where the topic is addressed with a variety of exercises.

Gender differences were evident in our findings: for example, female participants were more likely to react positively to the virtual coach overall. It could be that women are more likely than men to give positive feedback in general. In ***Academic performance*** (5A) where the coach shared his unpleasant feelings regarding pressure at school, female participants reacted to the content more negatively. It could be that girls are more likely to feel pressure about their grades and academic performance overall than boys [28] and therefore relate to the content differently. Similarly, the VC's utterances in ***Unpleasant thoughts*** (2A), were perceived more negatively by female participants. In a study concerning gender differences in connotative word processing, it was found that women tend to rate negative words more negatively than men, which could at least partially explain such results [8].

5.2 Directions for future research

Our findings indicate several areas of interest for future research and development, in which more evidence is needed for the development of design guidelines for conversational interfaces for adolescent mental well-being. In Table 8, we present potential issues to consider when developing the conversational style of conversational interfaces for supporting adolescents' mental well-being and intervention engagement based on our findings and prior research.

5.3 Limitations

Our findings should be interpreted considering the limitations of the study. First, survey response rates varied from 50% to 76% which could mean biased results, as only the most conscientious students might have logged in to the game platform to complete the surveys. Moreover, the results may not be comparable between conversation scripts, since the participants were from different schools on the different days of the event and thus each participant only evaluated one or two scripts.

Participants completed the surveys as part of a game, having a short time to answer the questions. They were rewarded for collecting points for each completed task where some of the participants might have been eager to complete the survey as quickly as possible to accumulate points, not giving full attention and consideration to their responses.

Iversen et al. [24] found that rewards, endorsements and storytelling can motivate teenagers, but individual motivations vary. In our study, participants might have been extrinsically motivated by point accumulation, but also by the task itself that provided possibly interesting and stimulating content, and a possibility to make an impact on research. However, the short time to answer very likely influenced the depth of the qualitative evaluations, resulting in relatively few participants addressing conversational style or content in their responses.

It was not possible for the participants to see what kind of conversational content preceded a script or followed it, and what kind of other content an online intervention program would have had surrounding the conversation. However, we believe that the standalone, noninteractive scripts were sufficient for measuring user perceptions since our study focused on evaluating the conversational style. The conversational interface was designed to have predefined responses for the virtual coach as well as for the interlocutor. Hence, the scripts can be said to adequately represent possible real conversations with the virtual coach.

As the focus of this study was on evaluating conversational style as a part of the iterative development of a conversational interface, the conversation scripts for the study were not designed as complete interventional exercises, and they were not situated in the context of the online ACT program, but rather presented

Table 8: Issues to consider when developing conversational interfaces for supporting adolescent mental well-being.

Issue to consider	Description	Rationale
Showing interest by asking questions about the interlocutor's life	Certain actions such as asking questions and showing interest may add to the perception of empathy.	Participants found the virtual coach empathetic when it asked questions and showed interest in the interlocutor's affairs.
Appropriate use of emojis	Use of emoijs is likely connected to the context and may not always be appropriate.	The use of emojis evoked mixed feelings in our study. Previous research has similar results [23].
Avoiding overt inquisitiveness	Being too curious or making too many unnecessary and consecutive inquiries might seem inquisitive to some.	Some participants considered the VC inquisitive, or even aggressive when it asked many questions or sent several messages in a row.
Fine-tuning the amount of encouragement	Too much encouragement in the messages can backfire.	Encouraging messages were mostly positively regarded, but excessive encouragement was seen as "motherly" or even arrogant.
Whether the conversational agent should have "personal experiences"	Conversational agents are virtual entities that cannot have personal experiences. Thus, when designing personal experiences, they might need to be believable for an artificial agent to have.	The VC's personal experiences were seen as disingenuous by some participants, even though previous studies show mainly positive response [6, 31, 33].
Artificiality can be considered appropriate	It might be acceptable, sometimes even preferable, for a virtual coach to seem artificial instead of human-like.	In certain cases within our study, artificiality was perceived positively, a finding that is also supported by previous research. [21, 30]
Avoiding cliché-like expressions especially when providing encouragement	Instead of generic and overused phrases, it might be better to keep it personal and original.	Phrases perceived as generic and/or overused evoked negative reactions in some participants.
Using colloquial language but not trying to imitate young people's way of speaking	Using formal language can make the agent seem artificial. "Youth language" produced by an adult can be seen as artificial too.	Colloquial language was preferred by the participants in our study as well as in the KIT chatbot study by Beilharz et al. [5]
Gender differences	It might be beneficial to customize the conversational style for male and female users as they may react to it differently.	E.g., female participants paid more attention to grammar mistakes. Some content divided opinions between male and female participants, such as in scripts 2A and 5A.
Being concise with message content especially when discussing more sensitive topics	Users seem to prefer concise messages.	Lengthy content chunks were criticized by participants in our study as well as in the KIT chatbot study by Beilharz et al. [5]

as stand-alone scripts. Thus, the generalizability of the findings is limited, and the findings should be interpreted in the general context of the design of engaging conversational interfaces for adolescents' mental well-being.

Qualitative evaluations might have been influenced by the semantic differential dimensions that preceded those evaluations as many participants used the same words to describe their impressions of the virtual coach.

6 CONCLUSION AND FUTURE WORK

This study contributes to the knowledge of conversational interface design for enhancing adolescents' intervention engagement as a part of a preventative approach to mental health. We conducted a quantitative analysis of the participants' evaluations of the conversational style used by the virtual coach in a selection of conversation scripts under development from the Youth Compass program. We also did a qualitative content analysis of participants' evaluations to reach a fuller understanding of their expectations. The overall response toward the virtual coach was positive, but we also discovered interesting themes regarding the conversational style that would warrant future research, including gender and personality-related differences in user preferences. It is known from prior research that users prefer computer systems that match their individual personality traits, such as extroversion and introversion [41]. Our results also suggest individual differences in conversational style preferences, so personality-adaptive conversational agents might be the

best way forward, as previous research has implicated [2, 42]. The results of this study and the proposed directions for future research can be utilized when planning further studies and designing the conversational style of conversational interfaces for adolescents, especially in the context of mental well-being.

ACKNOWLEDGMENTS

This study was supported by the Academy of Finland (No. 324638). We thank the city of Tampere and the participants of the study.

REFERENCES

[1] Alaa Ali Abd-Alrazaq, Asma Rababeh, Mohannad Alajlani, Bridgette M. Bewick, and Mowafa Househ. 2020. Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research* 22, 7 (July 2020), e16021. DOI:https://doi.org/10.2196/16021

[2] Rangina Ahmad, Dominik Siemon, Ulrich Gnewuch, and Susanne Robra-Bissantz. 2022. Designing Personality-Adaptive Conversational Agents for Mental Health Care. *Inf Syst Front* 24, 3 (June 2022), 923–943. DOI:https://doi.org/10.1007/s10796-022-10254-9

[3] Aino Ahtinen and Kirsikka Kaipainen. 2020. Learning and Teaching Experiences with a Persuasive Social Robot in Primary School – Findings and Implications from a 4-Month Field Study. In *Persuasive Technology. Designing for Future Change* (Lecture Notes in Computer Science), Springer International Publishing, Cham, 73–84. DOI:https://doi.org/10.1007/978-3-030-45712-9_6

[4] Petter Bae Bae Brandtzæg, Marita Skjuve, Kim Kristoffer Kristoffer Dysthe, and Asbjørn Følstad. 2021. When the Social Becomes Non-Human: Young People's Perception of Social Support in Chatbots. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (CHI '21), Association for Computing Machinery, New York, NY, USA, 1–13. DOI:https://doi.org/10.1145/3411764.3445318

[5] Francesca Beilharz, Suku Sukunesan, Susan L. Rossell, Jayashri Kulkarni, and Gemma Sharp. 2021. Development of a Positive Body Image Chatbot (KIT) With Young People and Parents/Carers: Qualitative Focus Group Study. *Journal of Medical Internet Research* 23, 6 (June 2021), e27807. DOI:https://doi.org/10.2196/27807

[6] Timothy Bickmore, Daniel Schulman, and Langxuan Yin. 2010. Maintaining Engagement in Long-term Interventions with Relational Agents. *Appl Artif Intell* 24, 6 (July 2010), 648–666. DOI:https://doi.org/10.1080/08839514.2010.492259

[7] Lorainne Tudor Car, Dhakshenya Ardhithy Dhinagaran, Bhone Myint Kyaw, Tobias Kowatsch, Shafiq Joty, Yin-Leng Theng, and Rifat Atun. 2020. Conversational Agents in Health Care: Scoping Review and Conceptual Analysis. *Journal of Medical Internet Research* 22, 8 (August 2020), e17158. DOI:https://doi.org/10.2196/17158

[8] Robert M. Chapman, Margaret N. Gardner, and Megan Lyons. 2022. Gender differences in emotional connotative meaning of words measured by Osgood's semantic differential techniques in young adults. *Humanit Soc Sci Commun* 9, 1 (April 2022), 119. DOI:https://doi.org/10.1057/s41599-022-01126-3

[9] Rachel G Curtis, Bethany Bartel, Ty Ferguson, Henry T Blake, Celine Northcott, Rosa Virgara, and Carol A Maher. 2021. Improving User Experience of Virtual Health Assistants: Scoping Review. *J Med Internet Res* 23, 12 (December 2021), e31737. DOI:https://doi.org/10.2196/31737

[10] Ronald E. Dahl, Nicholas B. Allen, Linda Wilbrecht, and Ahna Ballonoff Suleiman. 2018. Importance of investing in adolescence from a developmental science perspective. *Nature* 554, 7693 (February 2018), 441–450. DOI:https://doi.org/10.1038/nature25770

[11] Gilly Dosovitsky and Eduardo Bunge. 2023. Development of a chatbot for depression: adolescent perceptions and recommendations. *Child Adolesc Ment Health* 28, 1 (February 2023), 124–127. DOI:https://doi.org/10.1111/camh.12627

[12] Gilly Dosovitsky, Blanca S. Pineda, Nicholas C. Jacobson, Cyrus Chang, and Eduardo L. Bunge. 2020. Artificial Intelligence Chatbot for Depression: Descriptive Study of Usage. *JMIR Formative Research* 4, 11 (November 2020), e17065. DOI:https://doi.org/10.2196/17065

[13] Satu Elo and Helvi Kyngäs. 2008. The qualitative content analysis process. *J Adv Nurs* 62, 1 (2008), 107–115. DOI:https://doi.org/10.1111/j.1365-2648.2007.04569.x

[14] Ahmed Fadhil, Gianluca Schiavo, Yunlong Wang, and Bereket A. Yilma. 2018. The Effect of Emojis when interacting with Conversational Interface Assisted Health Coaching System. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare* (PervasiveHealth '18), Association for Computing Machinery, New York, NY, USA, 378–383. DOI:https://doi.org/10.1145/3240925.3240965

[15] Shuanghu Fang and Dongyan Ding. 2020. A meta-analysis of the efficacy of acceptance and commitment therapy for children. *Journal of Contextual Behavioral Science* 15, (January 2020), 225–234. DOI:https://doi.org/10.1016/j.jcbs.2020.01.007

[16] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment Health* 4, 2 (June 2017), e19. DOI:https://doi.org/10.2196/mental.7785

[17] Asbjørn Følstad and Petter Bae Brandtzaeg. 2020. Users' experiences with chatbots: findings from a questionnaire study. *Qual User Exp* 5, 1 (April 2020), 3. DOI:https://doi.org/10.1007/s41233-020-00033-2

[18] Silvia Gabrielli, Silvia Rizzi, Sara Carbone, and Valeria Donisi. 2020. A Chatbot-Based Coaching Intervention for Adolescents to Promote Life Skills: Pilot Study. *JMIR Human Factors* 7, 1 (February 2020), e16762. DOI:https://doi.org/10.2196/16762

[19] Hannah Gaffney, Warren Mansell, and Sara Tai. 2019. Conversational Agents in the Treatment of Mental Health Problems: Mixed-Method Systematic Review. *JMIR Mental Health* 6, 10 (October 2019), e14166. DOI:https://doi.org/10.2196/14166

[20] Sandra Garrido, Chris Millington, Daniel Cheers, Katherine Boydell, Emery Schubert, Tanya Meade, and Quang Vinh Nguyen. 2019. What Works and What Doesn't Work? A Systematic Review of Digital Mental Health Interventions for Depression and Anxiety in Young People. *Frontiers in Psychiatry* 10, (2019). Retrieved February 17, 2023 from https://www.frontiersin.org/articles/10.3389/fpsyt.2019.00759

[21] Stephanie Greer, Danielle Ramo, Yin-Juei Chang, Michael Fu, Judith Moskowitz, and Jana Haritatos. 2019. Use of the Chatbot "Vivibot" to Deliver Positive Psychology Skills and Promote Well-Being Among Young People After Cancer Treatment: Randomized Controlled Feasibility Trial. *JMIR Mhealth Uhealth* 7, 10 (October 2019), e15018. DOI:https://doi.org/10.2196/15018

[22] Rebecca Grist and Kate Cavanagh. 2013. Computerised Cognitive Behavioural Therapy for Common Mental Health Disorders, What Works, for Whom Under What Circumstances? A Systematic Review and Meta-analysis. *J Contemp Psychother* 43, 4 (December 2013), 243–251. DOI:https://doi.org/10.1007/s10879-013-9243-y

[23] Christine Grové. 2021. Co-developing a Mental Health and Wellbeing Chatbot With and for Young People. *Front Psychiatry* 11, (February 2021), 606041. DOI:https://doi.org/10.3389/fpsyt.2020.606041

[24] Elin Irene Krog Hansen and Ole Sejer Iversen. 2013. You are the real experts! Studying teenagers' motivation in participatory design. In *Proceedings of the 12th International Conference on Interaction Design and Children* (IDC '13), Association for Computing Machinery, New York, NY, USA, 328–331. DOI:https://doi.org/10.1145/2485760.2485826

[25] Steven C. Hayes. 2004. Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. *Behavior Therapy* 35, 4 (September 2004), 639–665. DOI:https://doi.org/10.1016/S0005-7894(04)80013-3

[26] Steven C. Hayes, Jacqueline Pistorello, and Michael E. Levin. 2012. Acceptance and Commitment Therapy as a Unified Model of Behavior Change. *The Counseling Psychologist* 40, 7 (2012), 976–1002. DOI:https://doi.org/10.1177/0011000012460836

[27] Erik Hedman, Brjánn Ljótsson, and Nils Lindefors. 2012. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. *Expert Rev Pharmacoecon Outcomes Res* 12, 6 (December 2012), 745–764. DOI:https://doi.org/10.1586/erp.12.67

[28] Björn Högberg. 2021. Educational stressors and secular trends in school stress and mental health problems in adolescents. *Social Science & Medicine* 270, (February 2021), 113616. DOI:https://doi.org/10.1016/j.socscimed.2020.113616

[29] James S. House, Robert L. Kahn, J.D. Mcleod, and D. Williams. 1985. Measures and concepts of social support. In *Social Support and Health* (Cohen, S. & Syme, S.L.). Academic Press, New York, NY, 83–108.

[30] Junhan Kim, Yoojung Kim, Byungjoon Kim, Sukyung Yun, Minjoon Kim, and Joongseek Lee. 2018. Can a Machine Tend to Teenagers' Emotional Needs? A Study with Conversational Agents. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI EA '18), Association for Computing Machinery, New York, NY, USA, 1–6. DOI:https://doi.org/10.1145/3170427.3188548

[31] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20), Association for Computing Machinery, New York, NY, USA, 1–12. DOI:https://doi.org/10.1145/3313831.3376175

[32] Susanna Lehtimaki, Jana Martic, Brian Wahl, Katherine T. Foster, and Nina Schwalbe. 2021. Evidence on Digital Mental Health Interventions for Adolescents and Young People: Systematic Overview. *JMIR Mental Health* 8, 4 (April 2021), e25847. DOI:https://doi.org/10.2196/25847

[33] Gale M. Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37, (August 2014), 94–100. DOI:https://doi.org/10.1016/j.chb.2014.04.043

[34] Audrey Mariamo, Caroline Elizabeth Temcheff, Pierre-Majorique Léger, Sylvain Senecal, and Marianne Alexandra Lau. 2021. Emotional Reactions and Likelihood of Response to Questions Designed for a Mental Health Chatbot Among Adolescents: Experimental Study. *JMIR Human Factors* 8, 1 (March 2021), e24343. DOI:https://doi.org/10.2196/24343

[35] Michael F. McTear. 2017. The Rise of the Conversational Interface: A New Kid on the Block? In *Future and Emerging Trends in Language Technology. Machine Learning and Big Data* (Lecture Notes in Computer Science), Springer International Publishing, Cham, 38–49. DOI:https://doi.org/10.1007/978-3-319-69365-1_3

[36] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine* 19, 2 (June 2012), 98–100. DOI:https://doi.org/10.1109/MRA.2012.2192811

[37] Robert R. Morris, Kareem Kouddous, Rohan Kshirsagar, and Stephen M. Schueller. 2018. Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions. *J Med Internet Res* 20, 6 (June 2018), e10148. DOI:https://doi.org/10.2196/10148

[38] Navot Naor, Alex Frenkel, and Mirène Winsberg. 2022. Improving Well-being With a Mobile Artificial Intelligence-Powered Acceptance Commitment Therapy Tool: Pragmatic Retrospective Study. *JMIR Form Res* 6, 7 (July 2022), e36018. DOI:https://doi.org/10.2196/36018

[39] Clifford Nass and Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. *Journal of social issues* 56, 1 (2000), 81–103. DOI:https://doi.org/10.1111/0022-4537.00153

[40] Rifat Rahman, Md. Rishadur Rahman, Nafis Irtiza Tripto, Mohammed Eunus Ali, Sajid Hasan Apon, and Rifat Shahriyar. 2021. AdolescentBot: Understanding Opportunities for Chatbots in Combating Adolescent Sexual and Reproductive Health Problems in Bangladesh. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (CHI '21), Association for Computing Machinery, New York, NY, USA, 1–15. DOI:https://doi.org/10.1145/3411764.3445694

[41] Byron Reeves and Clifford Nass. 1996. The media equation: how people treat computers, television, and new media like real people and places. CSLI Publications.

[42] Elayne Ruane, Sinead Farrell, and Anthony Ventresque. 2021. User Perception of Text-Based Chatbot Personality. In *Chatbot Research and Design* (Lecture Notes in Computer Science), Springer International Publishing, Cham, 32–47. DOI:https://doi.org/10.1007/978-3-030-68288-0_3

[43] Mark R. Scholten, Saskia M. Kelders, and Julia EWC Van Gemert-Pijnen. 2017. Self-Guided Web-Based Interventions: Scoping Review on User Needs and the Potential of Embodied Conversational Agents to Address Them. *Journal of Medical Internet Research* 19, 11 (November 2017), e7351. DOI:https://doi.org/10.2196/jmir.7351

[44] Mark R. Scholten, Saskia M. Kelders, and Julia E. W. C. Van Gemert-Pijnen. 2019. An Empirical Study of a Pedagogical Agent as an Adjunct to an eHealth Self-Management Intervention: What Modalities Does It Need to Successfully Support and Motivate Users? *Frontiers in Psychology* 10, (2019). Retrieved February 17, 2023 from https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01063

[45] Marita Skjuve, Ida Maria Haugstveit, Asbjørn Følstad, and Petter Brandtzaeg. 2019. Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human–chatbot interaction. (February 2019), 30–54. DOI:https://doi.org/10.17011/ht/urn.201902201607

[46] James G. Snider and Charles E. Osgood. 1969. *Semantic Differential Technique*. Aldine, Chicago.

[47] Marco Solmi, Joaquim Radua, Miriam Olivola, Enrico Croce, Livia Soardo, Gonzalo Salazar de Pablo, Jae Il Shin, James B. Kirkbride, Peter Jones, Jae Han Kim, Jong Yeob Kim, Andrè F. Carvalho, Mary V. Seeman, Christoph U. Correll, and Paolo Fusar-Poli. 2022. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Molecular psychiatry* 27, 1 (2022), 281–295. DOI:https://doi.org/10.1038/s41380-021-01161-7

[48] Taylor N Stephens, Angela Joerin, Michiel Rauws, and Lloyd N Werk. 2019. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. *Translational Behavioral Medicine* 9, 3 (June 2019), 440–447. DOI:https://doi.org/10.1093/tbm/ibz043

[49] Aditya Nrusimha Vaidyam, Hannah Wisniewski, John David Halamka, Matcheri S. Kashavan, and John Blake Torous. 2019. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *Can J Psychiatry* 64, 7 (2019), 456–464. DOI:https://doi.org/10.1177/0706743719828977

[50] Dana Vertsberger, Navot Naor, and Mirène Winsberg. 2022. Adolescents' Well-being While Using a Mobile Artificial Intelligence–Powered Acceptance Commitment Therapy Tool: Evidence From a Longitudinal Study. *JMIR AI* 1, 1 (November 2022), e38171. DOI:https://doi.org/10.2196/38171

[51] Xi Yang, Marco Aurisicchio, and Weston Baxter. 2019. Understanding Affective Experiences with Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), Association for Computing Machinery, New York, NY, USA, 1–12. DOI:https://doi.org/10.1145/3290605.3300772

[52] Xiaoyun Zhou, Sisira Edirippulige, Xuejun Bai, and Matthew Bambling. 2021. Are online mental health interventions for youth effective? A systematic review. *J Telemed Telecare* 27, 10 (2021), 638–666. DOI:https://doi.org/10.1177/1357633X211047285

[53] IBM Corp. Released 2021. IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp. https://ibm.com/spss

[54] ATLAS.ti Scientific Software Development GmbH. 2022. ATLAS.ti 22 for Windows, Version 22.1.5.0. https://atlasti.com

[55] WHO 2021. Mental health of adolescents. Retrieved February 21, 2023 from https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health

[56] Wysa - Everyday Mental Health. Retrieved February 21, 2023 from https://www.wysa.com/

[57] Woebot Health. Retrieved February 21, 2023 from https://woebothealth.com/