Text-based Healthcare Chatbots Supporting Patient and Health Professional Teams: Preliminary Results of a Randomized Controlled Trial on Childhood Obesity

Tobias Kowatsch[1(✉)], Marcia Nißen[2], Chen-Hsuan Iris Shih[2], Dominik Rüegger[2], Dirk Volland[1], Andreas Filler[1], Florian Künzler[2], Filipe Barata[2], Severin Haug[3], Dirk Büchter[4], Björn Brogle[4], Katrin Heldt[4], Pauline Gindrat[5], Nathalie Farpour-Lambert[6] & Dagmar l'Allemand[4]

[1] University of St. Gallen, Institute of Technology Management, St.Gallen, Switzerland
`tobias.kowatsch@unisg.ch`
[2] ETH Zurich, Department of Management, Technology and Economics, Zurich, Switzerland
[3] Swiss Research Institute for Public Health and Addiction, Zurich University, Zurich, Switzerland
[4] Children's Hospital of Eastern Switzerland, St. Gallen, Switzerland
[5] Fondation SportSmile, Nyon, Switzerland
[6] University Hospital of Geneva / University of Geneva, Department of Community Medicine, Primary Care and Emergency, Geneva, Switzerland

Abstract. Health professionals have limited resources and are not able to personally monitor and support patients in their everyday life. Against this background and due to the increasing number of self-service channels and digital health interventions, we investigate how text-based healthcare chatbots (THCB) can be designed to effectively support patients and health professionals in therapeutic settings beyond on-site consultations. We present an open source THCB system and how the THCP was designed for a childhood obesity intervention. Preliminary results with 15 patients indicate promising results with respect to intervention adherence (ca. 13.000 conversational turns over the course of 4 months or ca. 8 per day and patient), scalability of the THCB approach (ca. 99.5% of all conversational turns were THCB-driven) and over-average scores on perceived enjoyment and attachment bond between patient and THCB. Future work is discussed.

Keywords: Human-computer Interaction, Chatbot, Conversational Agent, Interpersonal Closeness, Attachment Bond, Counseling Psychology.

1   Introduction

Technology-based self-service channels [41] and digital health interventions [1, 31] have the potential to support patients in their everyday life and health professionals likewise. Although there are scalable self-service channels in the form of digital voice assistants and chatbots offered by Apple (Siri), Amazon (Alexa), Google (Assistant), Microsoft (Cortana) or Samsung (Bixby), they cannot (yet) be applied in healthcare

settings due to their lack of domain knowledge [32]. Thus, chatbots with a health focus, for example, Florence (getflorence.co.uk), Molly (sense.ly), Lark (lark.com), koko (itskoko.com) and various other messaging services, have recently gained interest in academia and industry with both inconclusive [20] and promising results related to user acceptance [5], working alliance [4] and treatment success [9, 14].

Text-based messaging services are "cheap, fast, democratic and popular" [15] and, especially for young people, the preferred way of communication [44]. We are therefore interested in effective designs of text-based healthcare chatbots (THCB) that focus on linguistic cues and a limited set of visual cues, i.e. usually only a small still image representing the agent, in contrast to embodied conversational agents discussed in prior work [4, 5, 8, 10, 11, 34]. And indeed, the efficacy of THCB approaches has already been shown [see 43 for an overview]. It is, however, open how to design THCB that support both patients and health professionals in therapeutic settings beyond on-site consultations as depicted in Fig. 1 and Fig. 2. We see THCBs as digital health coaches that use not only self-reports to infer the health condition of patients but also sensor information from everyday objects (e.g. from a smartphone [21], a car [19] or a PC mouse [16, 24-26, 42, 45]) and medical devices (e.g. a blood glucose meter [39]).
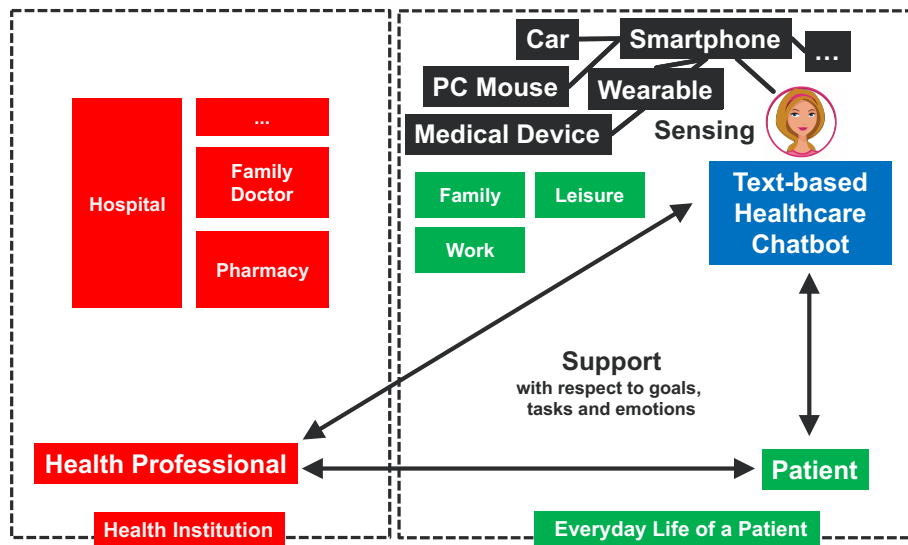

Fig. 1. A text-based healthcare chatbot using health data from everyday objects and self-reports from patients (sensing) to guide both patients and health professionals (support).

To address our research question, we propose and describe the design of a THCB and provide preliminary results of a clinical study as a part of an ongoing randomized controlled trial (RCT), in which the THCB was used to support not only young patients with obesity in reaching their daily intervention goals but also to support health professionals that were responsible for these teenagers.

The remainder of this paper is structured as follows. Next, we describe the open source behavioral intervention platform MobileCoach for THCBs and how it supports the

communication between patients and health professionals. We then describe the RCT and report first empirical results with respect to intervention adherence of the patient, scalability of the THCB, perceived enjoyment and attachment bond, a relationship quality of the working alliance inventory [17], between patient and THCB. We conclude this work with an outlook of future research.

2   Text-based Healthcare Chatbot

The THCB is part of the open source behavioral intervention platform MobileCoach www.mobile-coach.eu [12, 27]. It has already been evaluated in the public health context [14, 37] and provides a modular architecture and rule engine for the design of fully-automated digital health interventions. It also supports the implementation of RCTs and micro-randomized trials [22]. A mobile chat app has been recently introduced as a new chat client for MobileCoach [27]. An overview of the chat app's user interface is provided in Fig. 2.
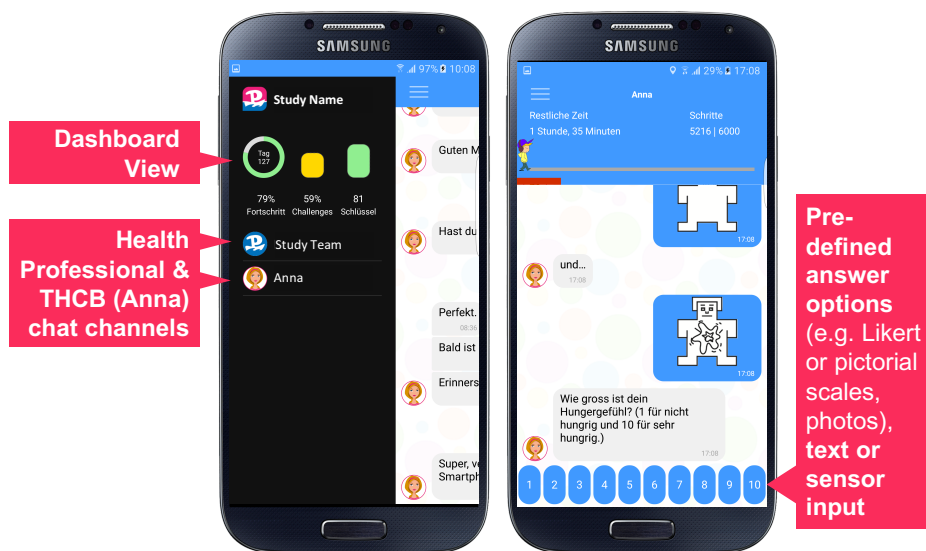

Fig. 2. Overview of the chat app's user interface. Note: Text-based Healthcare Chatbot (THCB)

The app allows the integration of visual THCB cues in a dedicated chat channel. This channel also provides pre-defined answer options for efficient chat interactions compared to traditional text-messaging systems. Moreover, we have implemented a second chat channel for patient and health professional communication like WhatsApp or iMessage and situations in which the THCB is unable to support patients in an automated way. For example, this channel can be used if a physician wants to motivate a patient in addition to the THCB or to ask patients to perform an ad-hoc task instead of missed intervention tasks such as "Dear John, why not go outside, run around about

5000 steps, make a selfie afterwards and send the picture back to me?" or because it was snowing the last night such as "Dear John, why not go outside, make a snowman, take a picture of it and send it back to me?" With that chat channel, patients are also able to get in direct contact with their health professionals, for example, to ask for help regarding an intervention task. Finally, MobileCoach uses a rule-based system to allow the THCB to automatically send notifications to health professionals or other individuals supporting the intervention like parents, sisters or peers of the patient. For example, if quality of life scores or other critical health states show a clear negative trend, which has been sensed either via chat-based answers fed back to the THCB or via smartphone sensors (e.g. no physical activity during the last 5 days), then the THCB can inform the physician via e-mail or SMS about that event. That is, health professionals do not have to actively monitor the conversational turns between the THCB and their patients but can design rules that trigger notifications that are relevant to them in an automated fashion. Finally, a dashboard view is integrated into the app to indicate general and automated feedback with respect to a digital health intervention, for example, to indicate the intervention progress, the number of goals achieved, average steps made or points earned.

3   Preliminary Evaluation

Against the background of the previous two sections, we now describe how a concrete THCB-based intervention for obese teenagers has been implemented with MobileCoach and present preliminary empirical results.

3.1   Design of the Text-based Healthcare Chatbot

The first THCB implementation based on MobileCoach and the novel chat app was collaboratively designed by computer scientists, physicians, a psychotherapist, diet and sport experts for a technology-supported intervention targeting childhood obesity. Linguistic and visual THCB characteristics were informed by the assumption that interpersonal closeness is positively related to attachment bond between patient and THCB [2, 17, 20, 38]. We therefore framed the THCB to represent a peer of the patient instead of an abstract entity such as the Google Assistant. There is a female and male version of the THCB, named Anna and Lukas, respectively. To clearly communicate the artificial character of the THCB and thus, not tricking patients in any way that they may interact with a real person, a comic profile image of an ordinary-looking teenager was used [33, 40] as depicted in Fig. 1 and Fig. 2 for the female version of the TCBH. For both task-related and social-interaction-related talk in terms of verbal cues [36], we used appropriate informal greetings and farewells [7, 28], the first name of the patient as the form of address [28], lay language [46] and the personal "Du" as T-V-distinction [29] used in German-speaking countries. Additionally, we used emoticons as quasi-nonverbal cues [47] and empathic feedback as relational cues [6, 23, 30]. The THCB also engaged in small-talk [3] on a non-regular basis and expressed happiness to see the patient and chat with him or her [4, 35].

3.2   Study Design, Intervention Details and Evaluation Measures

After a first pretest with 11 obese children in 2016, in which the THCB was positively evaluated [27] and the approval by the cantonal ethics committee, the study was started in January 2017. The study is still ongoing and has the objective to assess the efficacy of the THCB-based 5.5-month intervention compared to a treatment-as-usual control group without THCB support. In this 1-year trial, patients of the THCB group see their physicians four times during the intervention: at baseline, twice during the intervention and after the 5.5-month intervention. In addition, there are 2 telephone contacts during the intervention and follow up visits at 9 and 12 months. The primary medical outcome is the reduction of the sex- and age-adjusted body mass index standard deviation score (BMI-SDS) at the 1-year follow-up. Depending on the degree of challenge and therapy achievement, 1 out of 3 patients of both groups could win a smartphone after the intervention.

The THCB was introduced to the patients of the intervention group as an artificial assistant at baseline by the physicians and a dedicated study smartphone, a Samsung S6 with the chat app pre-installed, was handed out to the patients for the duration of the 5.5-month intervention. Physicians explained to their patients that they could choose between Anna or Lukas and that the corresponding THCB will come up with a challenge every day, i.e. the active intervention ingredients like doing a relaxation exercise (stress management module), counting steps as daily goal (physical activity module), taking photos of meals (diet module) or answering entertaining quiz questions (health literacy and entertainment module). Moreover, patients were told that the physicians could monitor the conversational turns between them and the THCB and that the results of the challenges would be reviewed in the next consultation hour.

Among other assessment instruments, we measured intervention adherence by the number of conversational turns per day in the chat app and the percentage of challenges that have been successfully completed. Scalability of the THCB was measured by the ratio of conversational turns in the manual chat channel vs. conversational turns in the THCB channel. Finally, we adopted the short version [13] of the attachment bond scale of the working alliance inventory [17] to measure the emotional and social relationship between patient and THCB. We further adopted one item of the perceived enjoyment scale [18] which was anchored on a 7-point Likert scale ranging from strongly disagree (1) to strongly agree (7). Intervention adherence and scalability were measured during the time of the intervention while the self-report instruments were employed after the 5.5-month intervention.

3.3   Interim Results of the THCB-based Intervention Group

We now present interim THCB-based results of the intervention group. For the adherence and scalability measures we report results of the first 4 months and for 15 patients (Age$_{mean}$ = 14.2, SD = 2.6, range 11.9 – 17.0). During the intervention 2 patients dropped out for medical reasons and did not fill out the questionnaire at the end of the intervention.

Results on intervention adherence are depicted in Fig. 3 and indicate that after 4 months, almost 70% of the patients had at least 4 conversational turns with the THCB

per day, i.e. the average number of turns to accept a challenge. Moreover, almost 40% of the daily challenges are completed successfully in month 4.
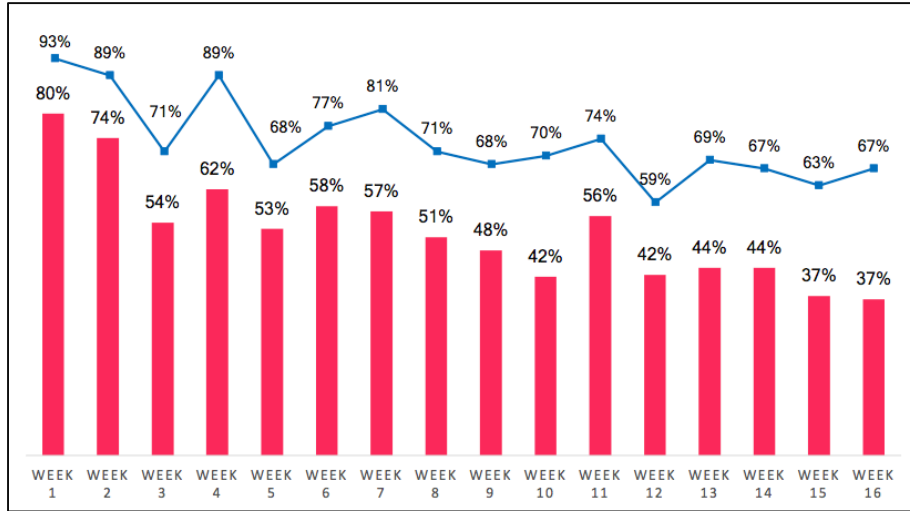

Fig. 3. Percentage of 15 patients with at least 4 conversational turns with the THCB per day on average (blue line chart on top) and percentage of successfully completed challenges (red bar chart on the bottom).

Results related to the scalability of the THCB are shown in Fig 4. Overall, 12.994 conversational turns have been recorded over the first 4 months and 15 patients. This results in approximately 8 conversational turns per day and patient.
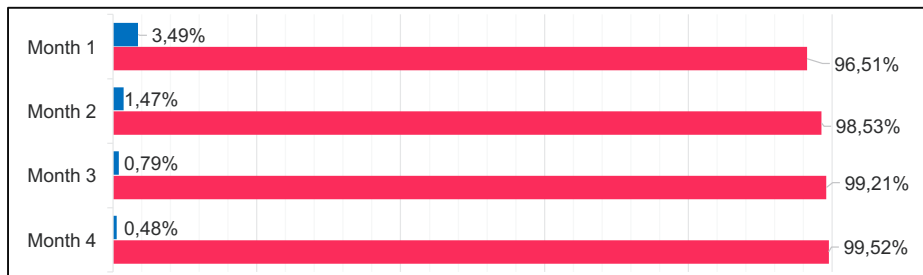

Fig. 4. Distribution of 12.994 conversational turns between patient and (a) health professionals (blue, top) and (b) THCB (pink, bottom) for the first four months of the intervention.

But only in the first month, 3.4% of the conversational turns took place in the manual chat channel with the study team and physicians. That was mainly because of technical issues and questions that were discussed over the first couple of weeks. As a result, less than 0.5% of all conversational turns took place in the manual chat channel in the fourth month indicating that the monitoring and support of the patients in their everyday life was mainly driven by interactions with the THCB. The health professionals were only

informed in cases, when there was no interaction between THCB and patients for three consecutive days, for example, when patients did not use the study smartphone for a longer period of time or were on holidays abroad and had no Internet access. On average, these notifications were triggered 4.1 times per patient and month (SD = 10.7).

Finally, the descriptive statistics related to perceived enjoyment and attachment bond are listed in Table 1. The 13 remaining patients indicated, that they enjoyed the THCB chat with Lukas and Anna. Moreover, attachment bond of the working alliance inventory between patients and the THCBs indicates also a high degree of social and emotional relationship at the end of the intervention.

Table 1. Descriptive statistics for perceived enjoyment and attachment bond (N=13). Note: Alpha = Cronbach's Alpha; Perceived enjoyment was anchored from strongly disagree (1) to strongly agree (7), attachment bond items from never (1) to always (7)

Construct	Items (Alpha)	Item / Example	Mean	SD
Perceived enjoyment	1 (N/A)	I enjoyed chatting with ___.	5.54	1.56
Attachment bond	4 (0.67)	I believe ___ liked me.	5.50	1.03

4 Summary and Future Work

In this work, we described one concrete instance of a text-based healthcare chatbot (THCB) system that was designed to support patients and health professionals likewise. Interim analysis of the intervention group from an ongoing RCT indicate that the implemented THCB, which took over the role of a peer character, engaged patients over four months to a remarkable extent. Moreover, more than 99.5% of the conversational turns were driven by the THCB which underlines their scalability of THCBs. Patients' perceptions regarding enjoyment and attachment bond with the THCB were also found to be good.

In our future work, we will assess additional measures (e.g. perceived interpersonal closeness between THCB and patient or incentive-based motivation) and medical outcomes of the RCT and compare them with the control group. We will also adapt the THCB to various other digital health interventions, for example, for individuals with asthma or diabetes or for substance abuse treatments. We are finally interested in investigating the health economic effects of THCBs in the context of various other non-communicable diseases to increase the efficiency of integrated care models and clinical pathways. Here, it is of utmost interest to identify ways to reallocate the limited resources of health professionals to those and only those patients that need personal, face-to-face care and where THCBs are likely to fail.

References

1. Agarwal, R., Gao, G., DesRoches, C., et al.: The Digital Transformation of Healthcare: Current Status and the Road Ahead. Information Systems Research 21, 796-809 (2010).
2. Aron, A., Aron, E.N., Smollan, D.: Inclusion of Other in the Self Scale and the structure of interpersonal closeness. Journal of Personality and Social Psychology 63, 596-612 (1992).
3. Bickmore, T., Cassell, J.: Social Dialogue with Embodied Conversational Agents. In: Kuppevelt, J.C.J., Bernsen, N.O., Dybkjær, L. (eds.) Advances in Natural Multimodal Dialogue Systems, vol. 30, pp. 23–54. Springer, Dordrecht (2005).
4. Bickmore, T., Gruber, A., Picard, R.: Establishing the computer–patient working alliance in automated health behavior change interventions. Patient Education and Counseling 59, 21-30 (2005).
5. Bickmore, T.W., Mitchell, S.E., Jack, B.W., et al.: Response to a relational agent by hospital patients with depressive symptoms. Interacting with Computers 22, 289-298 (2010).
6. Boukricha, H., Wachsmuth, I.: Modeling Empathy for a Virtual Human: How, When and to What Extent. The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3. International Foundation for Autonomous Agents and Multiagent Systems, 2011., pp. 1135–1136
7. Cassell, J., Bickmore, T.: Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents. User Modeling and User-Adapted Interaction 13, 89-132 (2003).
8. Cassell, J., Sullivan, J., Churchill, E.: Embodied Conversational Agents. MIT Press, Boston, MA (2000).
9. Cottrell, E., Chambers, R., O'Connell, P.: Using simple telehealth in primary care to reduce blood pressure: a service evaluation. BMJ Open 2, (2012).
10. Derrick, D.C., Jenkins, J.L., Nunamaker Jr., J.F.: Design Principles for Special Purpose, Embodied, Conversational Intelligence with Environmental Sensors (SPECIES) Agents. AIS Transactions on Human-Computer Interaction 3, 62-81 (2011).
11. Elkins, A.C., Derrick, D.C., Burgoon, J.K., et al.: Predicting Users' Perceived Trust in Embodied Conversational Agents Using Vocal Dynamics. 45th Hawaii International Conference on System Sciences (HICSS), pp. 579-588, Maui, HI (2012).
12. Filler, A., Kowatsch, T., Haug, S., et al.: MobileCoach: A Novel Open Source Platform for the Design of Evidence-based, Scalable and Low-Cost Behavioral Health Interventions - Overview and Preliminary Evaluation in the Public Health Context. 14th annual Wireless Telecommunications Symposium (WTS 2015). IEEE, USA, New York (2015).
13. Hatcher, R.L., Gillaspy, J.A.: Development and validation of a revised short version of the working alliance inventory. Psychotherapy Research 16, 12-25 (2006).
14. Haug, S., Paz, R., Kowatsch, T., et al.: Efficacy of a web- and text messaging-based intervention to reduce problem drinking in adolescents: Results of a cluster-randomised controlled trial. Journal of Consulting and Clinical Psychology 85, 147-159 (2017).
15. Herring, S.C.: Slouching Toward the Ordinary: Current Trends in Computer-Mediated Communication. New Media & Society 6, 26–36 (2004).
16. Hibbeln, M., Jenkins, J.L., Schneider, C., et al.: How Is Your User Feeling? Inferring Emotion Through Human-Computer interaction Devices. MIS Quarterly 41, 1-21 (2017).
17. Horvath, A.O., Greenberg, L.S.: Development and validation of the Working Alliance Inventory. Journal of Counseling Psychology 36, 223-233 (1989).
18. Kamis, A., Koufaris, M., Stern, T.: Using an Attribute-Based Decision Support System for User-Customized Products Online: An Experimental Investigation. MIS Quarterly 32, 159-177 (2008).

19. Kang, J.J., Adibi, S., Larkin, H., et al.: Predictive data mining for Converged Internet of Things: A Mobile Health perspective. International Telecommunication Networks and Applications Conference (ITNAC). IEEE, Sydney, Australia (2015).
20. Kiluk, B.D., Serafini, K., Frankforter, T., et al.: Only connect: The working alliance in computer-based cognitive behavioral therapy. Behaviour Research and Therapy 63, 139-146 (2014).
21. Kim, J., Lim, S., Lee, B., et al.: Detecting Depression of Cancer Patients with Daily Mental Health Logs from Mobile Applications. Thirty Sixth International Conference on Information Systems (ICIS) 2015, Forth Worth, Texas, USA (2015).
22. Klasnja, P., Hekler, E.B., Shiffman, S., et al.: Microrandomized Trials: An Experimental Design for Developing Just-in-Time Adaptive Interventions. Health Psychology 34, 1220-1228 (2015).
23. Klein, J., Moon, Y., Picard, R.W.: This computer responds to user frustration: Theory, design, and results. Interacting with Computers 14, 119–140 (2002).
24. Kowatsch, T., Wahle, F., Filler, A.: Design and Lab Experiment of a Stress Detection Service based on Mouse Movements. 11th Mediterranean Conference on Information Systems (MCIS). AIS, Genoa, Italy (2017).
25. Kowatsch, T., Wahle, F., Filler, A.: stressOUT: Design, Implementation and Evaluation of a Mouse-based Stress Management Service. In: Maedche, A., vom Brocke, J., Hevner, A. (eds.) Designing the Digital Transformation: DESRIST 2017 Research in Progress Proceedings, KIT Scientific Working Papers 64, pp. 37-45. KTI, Karlsruhe, Germany (2017).
26. Kowatsch, T., Wahle, F., Filler, A., et al.: Towards Short-Term Detection of Job Strain in Knowledge Workers with a Minimal-invasive Information System Service: Theoretical Foundation and Experimental Design. 23rd European Conference on Information Systems (ECIS), Münster, Germany (2015).
27. Kowatsch, T., Volland, D., Shih, I., et al.: Design and Evaluation of a Mobile Chat App for the Open Source Behavioral Health Intervention Platform MobileCoach. In: A., M., J., v.B., A., H. (eds.) Designing the Digital Transformation. DESRIST 2017. Lecture Notes in Computer Science, vol 10243, pp. 485-489. Springer, Berlin; Germany (2017).
28. Laver, J.: Linguistic routines and politeness in greeting and parting. Conversational routine 289304, (1981).
29. Levinson, S.C.: Pragmatics. Cambridge University Press (1983).
30. Lisetti, C., Amini, R., Yasavur, U., et al.: I Can Help You Change!: An Empathic Virtual Agent Delivers Behavior Change Health Interventions. ACM Transactions on Management Information Systems 4, 1–28 (2013).
31. Marsch, L., Lord, S., Dallery, J.: Behavioral Healthcare and Technology: Using Science-Based Innovations to Transform Practice. Oxford University Press, New York, USA (2014).
32. Miner, A.S., Milstein, A., Schueller, S., et al.: Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health. JAMA Intern Med. 176, 619-625 (2016).
33. Nowak, K.L., Rauh, C.: The Influence of the Avatar on Online Perceptions of Anthropomorphism, Androgyny, Credibility, Homophily, and Attraction. Journal of Computer-Mediated Communication 11, 153–178 (2005).
34. Nunamaker Jr., J.F., Derrick, D.C., Elkins, A.C., et al.: Embodied Conversational Agent-Based Kiosk for Automated Interviewing. Journal of Management Information Systems 28, 17-48 (2011).
35. Okun, B.F.: Effective helping: interviewing and counseling techniques. Brooks/Cole Pub, London, UK (1997).

36. Papangelis, A., Zhao, R., Cassell, J.: Towards a Computational Architecture of Dyadic Rapport Management for Virtual Agents. In: Bickmore, T., Marsella, S., Sidner, C. (eds.) Proc. 14th International Conference on Intelligent Virtual Agents, Boston, MA, vol. LNCS, volume 8637, pp. 320-324. Springer, London, UK (2014).
37. Paz, R., Haug, S., Kowatsch, T., et al.: Moderators of Outcome in a Technology-based Intervention to Prevent and Reduce Problem Drinking Among Adolescents. Addictive Behaviors 72, 64-71 (2017).
38. Popovic, M., Milne, D., Barrett, P.: The scale of perceived interpersonal closeness (PICS). Clinical Psychology & Psychotherapy 10, 286-301 (2003).
39. Quinn, C.C., Shardell, M.D., Terrin, M.L., et al.: Cluster-Randomized Trial of a Mobile Phone Personalized Behavioral Intervention for Blood Glucose Control. Diabetes Care 34, 1934-1942 (2011).
40. Sah, Y.J., Peng, W.: Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website. Computers in Human Behavior 45, 392-401 (2015).
41. Scherer, A., Wunderlich, N.V., von Wangenheim, F.: The Value of Self-Service: Long-Term Effects of Technology-Based Self-Service Usage on Customer Retention. MIS Quarterly 39, 177-200 (2015).
42. Seelye, A., Hagler, S., Mattek, N., et al.: Computer mouse movement patterns: A potential marker of mild cognitive impairment. Alzheimers Dement (Amst) 1, 472-480 (2015).
43. Silva, B.M.C., Rodrigues, J.J.P.C., La Torre Diez, I.d., et al.: Mobile-health: A review of current state in 2015. Journal of biomedical informatics 56, 265–272 (2015).
44. Smith, A., Page, D.: U.S. Smartphone Use in 2015. PewResearchCenter (2015).
45. Sun, D., Paredes, P., Canny, J.: MouStress: detecting stress from mouse motion. CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 61-70. ACM, Toronto, Canada (2014).
46. Vosbergen, S., Mulder-Wiggers, J., Lacroix, J.P., et al.: Using personas to tailor educational messages to the preferences of coronary heart disease patients. Journal of Biomedical Informatics 53, 100–112 (2015).
47. Walther, J.B., D'Addario, K.P.: The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication. Social Science Computer Review 19, 324–347 (2001).