




Inside pandora's box: a systematic review of the assessment of the perceived quality of chatbots for people with disabilities or special needs

Stefano Federici, Maria Laura de Filippis, Maria Laura Mele, Simone Borsci, Marco Bracalenti, Giancarlo Gaudino, Antonello Cocco, Massimo Amendola & Emilio Simonetti

To cite this article: Stefano Federici, Maria Laura de Filippis, Maria Laura Mele, Simone Borsci, Marco Bracalenti, Giancarlo Gaudino, Antonello Cocco, Massimo Amendola & Emilio Simonetti (2020): Inside pandora's box: a systematic review of the assessment of the perceived quality of chatbots for people with disabilities or special needs, Disability and Rehabilitation: Assistive Technology, DOI: [10.1080/17483107.2020.1775313](https://doi.org/10.1080/17483107.2020.1775313)

To link to this article: <https://doi.org/10.1080/17483107.2020.1775313>

 [View supplementary material](#) 

 Published online: 18 Jun 2020.

 [Submit your article to this journal](#) 

 Article views: 5






 [View related articles](#) 

 [View Crossmark data](#) 

REVIEW



Inside pandora's box: a systematic review of the assessment of the perceived quality of chatbots for people with disabilities or special needs

Stefano Federici^a , Maria Laura de Filippis^a , Maria Laura Mele^a , Simone Borsci^{b,c,d} , Marco Bracalenti^a , Giancarlo Gaudino^e, Antonello Cocco^e, Massimo Amendola^e and Emilio Simonetti^f

^aDepartment of Philosophy, Social and Human Sciences and Education, University of Perugia, Perugia, Italy; ^bDepartment of Cognitive Psychology and Ergonomics, Faculty of BMS, University of Twente, Enschede, The Netherlands; ^cDepartment of Surgery and Cancer, Faculty of Medicine, NIHR London IVD, Imperial College, London, UK; ^dDesign Research Group, School of Creative Arts, Hertfordshire University, Hatfield, UK; ^eDGTCSI-ISCTI – Directorate General for Management and Information and Communications Technology, Superior Institute of Communication and Information Technologies, Ministry of Economic Development, Rome, Italy; ^fDepartment of Public Service, Prime Minister's Office, Rome, Italy

ABSTRACT

Introduction: People with disabilities or special needs can benefit from AI-based conversational agents (i.e., chatbots) that are used for competence training and well-being management. Assessing the quality of interactions with these chatbots is key to being able to reduce dissatisfaction with them and to understanding their potential long-term benefit. This in turn will help to increase adherence to their use, thereby improving the quality of life of the large population of end-users that they are able to serve.

Methods: Following Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) methodology, we systematically reviewed the literature on methods of assessing the perceived quality of interactions with chatbots using the from Scopus and the Web of Science electronic databases. Using the Boolean operators (AND/OR) the keywords chatbot*, conversational agent*, special needs, disability were combined.

Results: Revealed that only 15 of 192 papers on this topic included people with disabilities or special needs in their assessments. The results also highlighted the lack of a shared theoretical framework for assessing the perceived quality of interactions with chatbots.

Conclusion: Systematic procedures based on reliable and valid methodologies continue to be needed in this field. The current lack of reliable tools and systematic methods to assess chatbots for people with disabilities and special needs is concerning, and ultimately, it may also lead to unreliable systems entering the market with disruptive consequences for people.

ARTICLE HISTORY

Received 27 March 2020
Accepted 25 May 2020

KEYWORDS

Chatbots; conversational agents; assistive technology; people with disability; people with special needs; quality of interaction; usability; user experience

► IMPLICATIONS FOR REHABILITATION

- Chatbots applied in rehabilitation are mainly tested in terms of clinical effectiveness and validity with a minimal focus on measuring the quality of the interaction
- The usability and interactive properties of chatbots applied in rehabilitation are not comparable as each tool is measured in different way
- The lack of a common framework to assess chatbots exposes people with disability and special needs to the risk of using unreliable tools



Introduction


Chatbots are intelligent conversational software that can interact with humans *via* text-based dialogue using natural language [1]. They are widely used to support people services, decision-making and training in various domains [2–5].

Despite some ethical and societal concerns around privacy and security [6,7], chatbots and AI-based conversational agents that have human avatars are becoming more common. Moreover, there is wide consensus on their usefulness, especially in the domain of health where they can support rehabilitation, adherence to treatment, and training [8,9]. People with disabilities or special needs can also benefit from AI support systems, in terms

of training in various competences and managing their well-being [3,10,11]. For this reason, assessing the perceived quality of interaction with chatbots is key to being able to reduce dissatisfaction and to understanding their potential long-term benefit, in order to increase adherence and thereby improve the quality of life of the large population of end-users that they are able to serve.

Chatbots are interaction systems. Thus, independent of their domain of application, their performance, in terms of the quality of that interaction, should be designed and assessed with the users with whom they interact, rather than through a system-centred approach [1]. A recent review by Abd-alrazaq et al. [8] showed that in the mental health domain, researchers usually

CONTACT Stefano Federici  stefano.federici@unipg.it  Department of Philosophy, Social and Human Sciences and Education, University of Perugia, Piazza G. Ermini, 1, Perugia, 06126, Italy

 Supplemental data for this article can be accessed [here](#).

© 2020 Informa UK Limited, trading as Taylor & Francis Group

assess chatbots by randomized controlled trial only. Quality of interaction is rarely assessed; where it is, it is done by looking at unstandardized aspects of interaction and qualitative scales that do not allow comparisons to be made. This inconsistent way of testing the quality of interaction of these devices or applications through a wide and varying range of factors is endemic to every domain that uses chatbots and makes it hard to benchmark the results of studies [1,12,13]. While some qualitative guidelines and tools are emerging [1,14], it is still hard to find an agreement on what factors should be tested. As argued by Park and Humphry [15], the development of these innovative systems should proceed around a common framework for assessing the perceived quality of interaction, in order to prevent chatbots being viewed by their end-users as simply another source of social exclusion and being abandoned as quickly as any other inconsistent assistive technology would be [16,17]. Therefore, a common framework and guidelines on how to assess the perceived quality of interaction of chatbots are needed.

From a systems perspective, the subjective experience of quality derives from the interaction between the user and the application under specific conditions and contexts. Subjective experience cannot be assessed simply by assuming that satisfactory performance of the system as perceived by the user equates to good user experience [18]. As summarized by Lewis [19], the need to measure objective and subjective aspects of interactions in a reliable and comparable way is a lesson that those in the field of the human-computer interaction have learnt; it has yet to be learnt in the field of chatbots, as also recently highlighted by Bendig et al. [20].

Because of the lack of a shared assessment framework defining comparable evaluation criteria, chatbot developers are forced to rely on the umbrella framework provided by International Organisation for Standards (ISO) 9241-11 [21] for assessing usability and by ISO 9241-210 [22] for assessing user experience (UX). These two ISO standards define the key factors of interaction quality: (i) effectiveness, efficiency and satisfaction in a specific context of use (ISO 9241-11); and (ii) the control (where possible) over time of expectations concerning use, satisfaction, perceived level of acceptability, trust, usefulness and all those factors that ultimately push users to adopt and keep using a tool (ISO 9241-210). Although these standards have not yet been adapted to accommodate the specific needs of chatbots and conversational agents, these two aspects – usability and UX – are essential to the perceived quality of interaction [23]. Until a framework has been developed and broad consensus on the assessment criteria established, practitioners may benefit from assessing chatbots against these ISO standards; this would allow them to compare the interactive performance of these applications.

This paper investigates how factors of perceived quality of interaction are measured in studies of AI-based agents that support people with disabilities or special needs. Our systematic literature review was performed in accordance with the PRISMA reporting checklist ([Supplemental material](#)).

Methods

Study design

This systematic review was performed on journal articles examining the interaction between chatbots and people with disabilities or special needs over the last 10 years (our eligibility criteria and electronic search strategy are described in [Supplemental material](#)).

Research questions

To investigate whether and how the quality of interaction with chatbots are evaluated in line with ISO standards of usability (ISO 9241-11) and user experience (ISO 9241-210), the review sought to answer the following research questions:

- RQ1 – How are key usability factors measured and reported in evaluations of chatbots for people with disabilities or special needs?
- RQ2 – How are factors relating to user experience measured and reported in assessments of chatbots?

Eligibility criteria

In the review, we included records that:

1. mentioned chatbots or conversational interfaces/agents for people with disabilities or special needs in the title, abstract, keywords or main text.
2. included empirical findings and discussion on theories (or frameworks) to do with factors that might contribute to the perceived quality of interaction with chatbots, with a focus on people with various types of disability.

We excluded records that did not include at least one group of end-users with a disability in either the testing or the design of the interaction, as well as those studies that focussed on:

1. testing emotion recognition during the interaction exchange, or assessing applications for detecting the development of disability conditions or disease.
2. chatbots supporting people with alcoholism, anxiety, depression or traumatic disorders.
3. the assessment of end-user compliance with clinical treatment, or assessing the clinical effectiveness of using AI agents as an alternative to standard (or other) forms of care without considering the interaction exchange with the chatbot.
4. ethical and legal implications of interacting with AI-based digital tools.

Search strategy

Records were retrieved from Scopus and the Web of Science using the Boolean operators (AND/OR) to combine the following keywords: Chatbot, Chatbots, conversational agent, conversational agents, special needs, disability. We searched only for English language articles, as reported in [Supplemental material](#).

Results

As shown in [Figure 1](#), a total of 147 items were retrieved from Scopus and the Web of Science. A further 53 records were added from a previous review of chatbots used in mental health by Abdalrazaq et al. [8]. After removing eight duplicates, a scan of the remaining 192 records by title and abstract was performed by two authors (MLDF, SB). Articles that mentioned their scope as being to assess interactions between chatbots and conversational agents and people with various types of intellectual disabilities or special needs were retained.

The full text of 68 records was then scanned to look for articles mentioning methods and factors for assessing the interactions of people with disabilities or special needs with chatbots. The final list comprised 15 records [3,10,11,25–36].

Of the 15 records that matched our criteria, 80% investigated AI agents for supporting people with autism and (mild to severe)

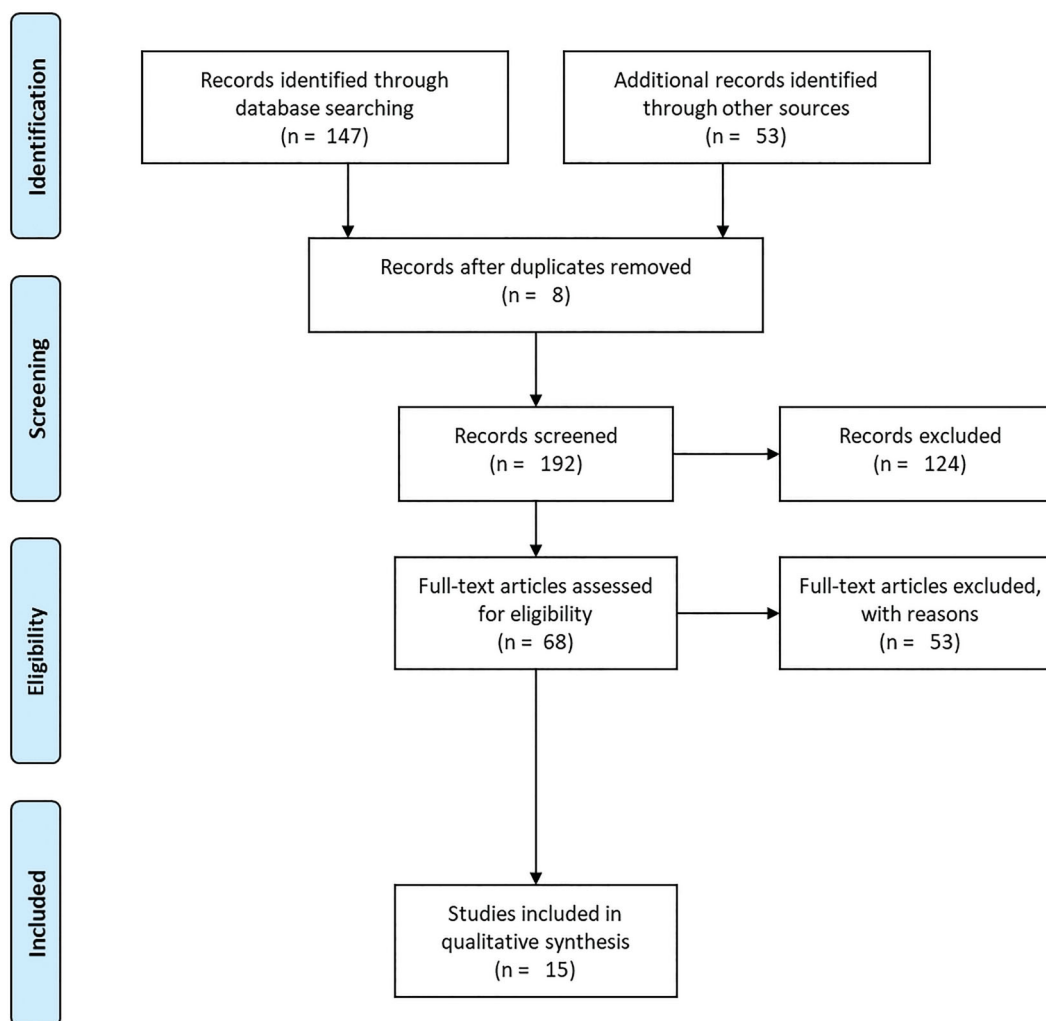


Figure 1. A pictorial view of the review process in accordance with the flowchart of the PRISMA guidelines [24].

Table 1. Overall goal of the chatbot and type of research approach used in each study.

Study number	Goal of the tools	Research approach
Ali et al. [25]	Health and rehabilitation	Survey/Questionnaire
Beaudry et al. [3]	Health and rehabilitation	Survey/Questionnaire
Burke et al. [10]	Health and rehabilitation	Quasi-experiment
Cameron et al. [26]	Health and rehabilitation	Survey/Questionnaire
Ellis et al. [11]	Health and rehabilitation	Quasi-experiment
Konstantinidis et al. [27]	Health and rehabilitation	Survey/Questionnaire
Lahiri et al. [28]	Health and rehabilitation	Survey/Questionnaire
Ly et al. [29]	Health and rehabilitation	Randomized controlled trials
Milne et al. [31]	Health and rehabilitation	Quasi-experiment
Razavi et al. [31]	Learning/education/training	Survey/Questionnaire
Smith et al. [32]	Learning/education/training	Randomized controlled trials
Smith et al. [33]	Learning/education/training	Randomized controlled trials
Smith et al. [34]	Learning/education/training	Randomized controlled trials
Tanaka et al. [35]	Learning/education/training	Quasi-experiment
Wargnier et al. [36]	Health and rehabilitation	Survey/Questionnaire

mental disabilities; the other 20% focussed on testing applications for supporting the general health or training of people with a wide range of disabilities. Table 1 shows that the main goal of 66.6% of the applications was to support health and rehabilitation; the remaining studies focussed on solutions to support learning and training for people with disabilities. In terms of their approach to assessment, 46.7% of the studies used surveys or questionnaires, 26.7% applied a quasi-experimental procedure

and the remaining 26.7% tested chatbots through randomized controlled trials that assessed some aspects relating to quality of interaction.

As shown in Table 2, factors relating to usability (i.e., effectiveness, efficiency and satisfaction) were partly assessed, with 80% of the studies reporting measures of effectiveness, 26.7% measures of efficiency and 20% measures of satisfaction. As for UX, acceptability was the most commonly reported measure (26.7% of the

Table 2. Main factors considered in assessing the quality of interaction in each study, and notes on methodological gaps (notes on methods).

Study ID	Usability factors			User Experience factors		Notes on methods
	Effectiveness	Efficiency	Satisfaction	Acceptability	Other factors	
1			X	X		<ul style="list-style-type: none"> Scale inspired by the System Usability Scale (SUS), and qualitative questions Items unknown
2	X					
3	X					
4	X		X			<ul style="list-style-type: none"> SUS plus a heuristic-based inspection tool
5	X	X	X	X	<ul style="list-style-type: none"> Safety as a lack of adverse events 	<ul style="list-style-type: none"> Not validated satisfaction Items unknown Assessment over a period of time Qualitative questions about functions and experience of interactions Items unknown
6				X		<ul style="list-style-type: none"> Acceptability inferred as lack of complaints Engagement inferred by data General survey about interaction (Items unknown)
7	X	X		X	<ul style="list-style-type: none"> Engagement 	<ul style="list-style-type: none"> Engagement inferred by data General survey about interaction (Items unknown)
8	X	X			<ul style="list-style-type: none"> Overall experience 	<ul style="list-style-type: none"> Phone interview Items unknown
9	X				<ul style="list-style-type: none"> Overall experience Intention to use 	<ul style="list-style-type: none"> Survey to inform redesign, not to assess Items unknown
10					<ul style="list-style-type: none"> Overall experience 	<ul style="list-style-type: none"> Items unknown
11	X				<ul style="list-style-type: none"> Helpfulness 	<ul style="list-style-type: none"> Items unknown
12	X				<ul style="list-style-type: none"> Ease of use Helpfulness Enjoyment 	<ul style="list-style-type: none"> Items unknown
13	X					
14	X					
15	X	X			<ul style="list-style-type: none"> Appearance experience 	<ul style="list-style-type: none"> Items unknown
Total (%)	80%	26.7%	20%	26.7%		

cases) while a few other factors (e.g., engagement, safety, helpfulness, etc.) were measured using various approaches.

Discussion

The results suggest that the main focus of studies of chatbots for people with disabilities or special needs is the effectiveness of such applications, compared with standard practice, in supporting adherence to treatment. From a larger sample of 68 published records, only 22% actually involved end-users in their assessments of the quality of interaction.

In line with our research questions, the results can be summarized as follows:

RQ1. A total of 80% of the studies [3,10,11,26,28,29,31,33–36] tested the effectiveness of chatbots according to the ISO standard [21], i.e., the ability of people to perform correctly and achieve their goals. Only 26.7% of the studies [11,28,29,36] also investigated efficiency, by measuring time taken to perform or factors relating to the resources invested by participants to achieve their goals. Only 20% [11,25,26] mentioned the intention of gathering data on user satisfaction in a structured way despite only one study [26] using a validated scale, while in one study practitioners adapted a standardized questionnaire without clarifying the changes in the items [25] and a qualitative scale was used in another [11].

RQ2. Acceptability was mentioned as an assessment factor in 26.7% of the studies [11,25,27,28]. Despite the popularity of the Technology Acceptance Model [37,38], acceptability was assessed in various ways – e.g., lack of complaints [28] – or as an alternative to satisfaction [27]. A total of 53% of the studies used various factors to assess quality of interaction, such as overall experience, safety, acceptability, engagement, intention to use, ease of use, helpfulness, enjoyment and appearance. Most used unstandardized questionnaires to assess quality

of interaction. Even when a factor such as safety was presented as a reasonable way of controlling quality, in accordance with ISO standards for assessing medical devices [39] and risk analysis [40], the method of its measurement in these studies was questionable i.e., assessing a product to be safe because of the lack of adverse events [11].

The results of the present study suggest that informal and untested measures of quality are often employed when it comes to evaluating user interactions with AI agents. This is particularly relevant in the health and well-being domain, where researchers set out to measure the clinical validity of tools intended to support people with disabilities or special needs. The risk is that shortcomings in these methods could significantly compromise the quality of chatbot usage, ultimately leading to the abandonment of applications that could otherwise have a positive impact on their end-users.

Two limitations of the present study should be highlighted. First, only two of the largest databases were included in the literature search (Scopus and the Web of Science). Future studies should expand this analysis by including other libraries, such as medical databases (e.g., PubMed) or more technologically oriented repositories such as IEEEExplore and the ACM digital library. Second, we focussed only on studies that involved end-users in the assessment of chatbots; future analyses might attempt to explore and define common criteria to predict and simulate interaction between conversational agents and people with disabilities or special needs.

Despite these limitations, the present work suggests that practitioners do not differentiate between clinical effectiveness (i.e., the optimum clinical result of using a device compared with standard practice) and interaction effectiveness (i.e., the ability of

end-users to efficiently perform the tasks required using the application). In the health field, effectiveness is often assessed by gathering evidence about the clinical validity of an application; minimal focus is given to measuring interaction effectiveness in order to assess the usability of chatbots.

The current literature also reveals that satisfaction, together with other subjective aspects relating to UX, is usually reduced to a set of qualitative questions generated by researchers to fit the needs of their evaluation procedure. While this suggests a concerning lack of knowledge around the methods and goals of interaction assessment, this is not a new issue; elsewhere, Borsci et al. [41] have already described the wide use of unreliable measures of satisfaction in health. Clinicians often focus on demonstrating the clinical effectiveness and usefulness of technological interventions; at the same time, they show minimal interest in using standardized approaches to evaluate subjective aspects of the interaction [41], despite these being considered key aspects that should be assessed as rigorously as are objective aspects of interaction (i.e., effectiveness and efficiency).

This lack of a wider perspective has resulted in a lack of comparable data that might be used to model the quality of human–chatbot interactions. As a consequence, the clinical effectiveness of every chatbot designed to deliver a health service is compared only against standard practice; they cannot be compared, in terms of interaction efficiency, effectiveness, satisfaction (i.e., usability) or user experience, against other applications intended to deliver the same or similar services.

Conclusion

AI interactive agents are going to shape and change our future daily lives, with a potentially positive impact on our well-being. Nevertheless, to fully model and maximize the impact of such applications it will be necessary to identify a common framework for their evaluation. The current lack of reliable tools and systematic methods of assessment is not only concerning; ultimately, it may also lead to unreliable systems entering the market with disruptive consequences for people. This review has enabled us to open the Pandora's box [42] of quality interactions with chatbots. Inside this box we have found a lack of standards and guidelines. Three following main concerns regarding the evaluation of chatbots for people with disabilities or special needs can be viewed as the takeaway points of this study:

- A lack of complaints or the occurrence of adverse events cannot be considered appropriate measures of safety or acceptability. Even in the absence of problems, the lack of an event cannot reliably indicate a lack of risk associated with using an application.
- Satisfaction and acceptability are two different constructs and should be measured separately.
- A set of qualitative questions tailored to an application is one way of gathering end-users' views and is certainly useful for informing any redesign required. However, as previously argued by Dillon [43], we need to cast the net much further, beyond the usual measures of quality user experience, in order to obtain all the data we need on the experiences of users. Moreover, the data must be reliable and collected in a way that enables the results to be replicated and compared. When it comes to user satisfaction or perceived usability and experience, such reliable measures do exist (for a review, see: [44]), but adapted and validated tools for assessing AI conversational agents remain lacking.

Disclosure statement

The authors report that they have no conflicts of interest.

Funding

This study was supported by DGTCSI-ISCTI – Directorate General for Management and Information and Communications Technology, Superior Institute of Communication and Information Technologies, Ministry of Economic Development, Rome, Italy, under grant Project “eGLU-box PRO” on October 6, 2019.

ORCID

Stefano Federici  <http://orcid.org/0000-0001-5681-0633>
 Maria Laura de Filippis  <http://orcid.org/0000-0003-3282-3514>
 Maria Laura Mele  <http://orcid.org/0000-0002-2714-4683>
 Simone Borsci  <http://orcid.org/0000-0002-3591-3577>
 Marco Bracalenti  <http://orcid.org/0000-0002-8768-3793>

References

- [1] Radziwill NM, Benton MC. Evaluating quality of chatbots and intelligent conversational agents. arXiv Preprint. 2017: 170404579.
- [2] Ammari T, Kaye J, Tsai JY, et al. Music, search, and IoT: how people (really) use voice assistants. *ACM Trans Comput-Hum Interact.* 2019;26:1–28.
- [3] Beaudry J, Consigli A, Clark C, et al. Getting ready for adult healthcare: designing a chatbot to coach adolescents with special health needs through the transitions of care. *J Pediatr Nurs.* 2019;49:85–91.
- [4] Costa S, Brunete A, Bae BC, et al. Emotional storytelling using virtual and robotic agents. *Int J Human Robot.* 2018; 15:1850006.
- [5] D'mello S, Graesser A. AutoTutor and affective autotutor: learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Trans Interact Intell Syst.* 2012;2:1–39.
- [6] Birhane A, van Dijk J. Robot rights? Let's talk about human welfare instead. arXiv Preprint. 2020:200105046.
- [7] Pasquale F. A rule of persons, not machines: the limits of legal automation. *Geo Wash L Rev.* 2019;87:1.
- [8] Abd-Alrazaq AA, Alajlani M, Alalwan AA, Bewick BM, et al. An overview of the features of chatbots in mental health: a scoping review. *Int J Med Inf.* 2019;132:103978.
- [9] Fadhil A, Wang Y, Reiterer H. Assistive conversational agent for health coaching: a validation study. *Methods Inf Med.* 2019;58:9–23.
- [10] Burke SL, Bresnahan T, Li T, et al. Using virtual interactive training agents (ViTA) with adults with autism and other developmental disabilities. *J Autism Dev Disord.* 2018;48: 905–912.
- [11] Ellis T, Latham NK, Deangelis TR, et al. Feasibility of a virtual exercise coach to promote walking in community-dwelling persons with Parkinson disease. *Am J Phys Med Rehabil.* 2013;92:472–485.
- [12] Balaji D, Borsci S. Assessing user satisfaction with information chatbots: a preliminary investigation. University of Twente; 2019.
- [13] Tariverdiyeva G, Borsci S. Chatbots' perceived usability in information retrieval tasks: an exploratory analysis. University of Twente; 2019.

- [14] IBM Conversational UX. Talk meets technology – conversation design guidelines. 2018 [cited 2018 May 2]. Available from: <http://conversational-ux.mybluemix.net/design/conversational-ux/practices/>
- [15] Park S, Humphry J. Exclusion by design: intersections of social, digital and data exclusion. *Inf Commun Soc.* 2019; 22:934–953.
- [16] Federici S, Borsci S. Providing assistive technology in Italy: the perceived delivery process quality as affecting abandonment. *Disabil Rehabil Assist Technol.* 2016;11:22–31.
- [17] Scherer MJ, Federici S. Why people use and don't use technologies: Introduction to the special issue on assistive technologies for cognition/cognitive support technologies. *NeuroRehabilitation.* 2015;37:315–319.
- [18] Bevan N. Measuring usability as quality of use. *Software Qual J.* 1995;4:115–130.
- [19] Lewis JR. Usability: lessons learned ... and yet to be learned. *Int J Hum Comput Interact.* 2014;30:663–684.
- [20] Bendig E, Erb B, Schulze-Thuesing L, et al. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health – a scoping review. *Verhaltenstherapie.* 2019.
- [21] ISO. ISO 9241-11:2018 Ergonomic requirements for office work with visual display terminals – part 11: guidance on usability. Brussels (Belgium): CEN; 2018.
- [22] ISO. ISO 9241-210:2010 Ergonomics of human-system interaction – part 210: human-centred design for interactive systems. Brussels (Belgium): CEN; 2010.
- [23] Borsci S, Federici S, Malizia A, et al. Shaking the usability tree: why usability is not a dead end, and a constructive way forward. *Behav Inform Technol.* 2019;38:519–532.
- [24] Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* 2009;6:e1000100.
- [25] Ali MR, Razavi Z, Mamun AA, et al. A virtual conversational agent for teens with autism: experimental results and design lessons. *arXiv Preprint.* 2018:181103046.
- [26] Cameron G, Cameron D, Megaw G, et al. Assessing the usability of a chatbot for mental health care. In: Bodrunova SS, Koltsova O, Følstad A, et al., editors. *International Conference on Internet Science, LNCS 11551.* Cham (Switzerland): Springer; 2018. p. 121–132.
- [27] Konstantinidis EI, Hitoglou-Antoniadou M, Luneski A, et al. Using affective avatars and rich multimedia content for education of children with autism. *Proceedings of the 2nd International Conference on PErvasive Technologies Related to Assistive Environments: PETRA '09;* 2009 Jun 9–13; New York, NY. 2009. p. 1–6.
- [28] Lahiri U, Bekele E, Dohrmann E, et al. Design of a virtual reality based adaptive response technology for children with autism. *IEEE Trans Neural Syst Rehabil Eng.* 2013;21: 55–64.
- [29] Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. *Internet Interv.* 2017;10:39–46.
- [30] Smith MJ, Fleming MF, Wright MA, et al. Job offers to individuals with severe mental illness after participation in virtual reality job interview training. *Psychiatr Serv.* 2015;66: 1173–1179.
- [31] Milne M, Luerssen MH, Lewis TW, et al. Development of a virtual agent based social tutor for children with autism spectrum disorders. *Proceedings of the International Joint Conference on Neural Networks: IJCNN 2010;* 2010 Jul 18–23; Barcelona, Spain. 2010. p. 1–9.
- [32] Razavi SZ, Ali MR, Smith TH, et al. The LISSA virtual human and ASD teens: an overview of initial experiments. In: Traum D, Swartout W, Khooshabeh P, et al., editors. *International Conference on Intelligent Virtual Agents.* Cham (Switzerland): Springer; 2016. p. 460–463.
- [33] Smith MJ, Ginger EJ, Wright K, et al. Virtual reality job interview training in adults with autism spectrum disorder. *J Autism Dev Disord.* 2014;44:2450–2463.
- [34] Smith MJ, Ginger EJ, Wright MA, et al. Virtual reality job interview training for individuals with psychiatric disabilities. *J Nerv Ment Dis.* 2014;202:659–667.
- [35] Tanaka H, Negoro H, Iwasaka H, et al. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PLoS ONE.* 2017;12:e0182151.
- [36] Wargnier P, Benveniste S, Jouvelot P, et al. Usability assessment of interaction management support in LOUISE, an ECA-based user interface for elders with cognitive impairment. *TAD.* 2018;30:105–126.
- [37] Venkatesh V, Morris MG, Davis GB, et al. User acceptance of information technology: toward a unified view. *MIS Q.* 2003;27:425–478.
- [38] Federici S, Tiberio L, Scherer MJ. Ambient assistive technology for people with dementia: an answer to the epidemiologic transition. In: Combs D, editor. *New research on assistive technologies: uses and limitations.* New York (NY): Nova Publishers; 2014. p. 1–30.
- [39] IEC. IEC 62366-1:2015 medical devices – part 1: application of usability engineering to medical devices. Brussels (Belgium): CEN; 2015.
- [40] ISO. ISO 14971:2007 medical devices – application of risk management to medical devices. Brussels (Belgium): CEN; 2007.
- [41] Borsci S, Buckle P, Walne S. Is the LITE version of the usability metric for user experience (UMUX-LITE) a reliable tool to support rapid assessment of new healthcare technology? *Appl Ergon.* 2020;84:103007.
- [42] Wikipedia contributors. Pandora's box. *Wikipedia, the Free Encyclopedia;* 2020.
- [43] Dillon A. Beyond usability: process, outcome and affect in human computer interactions. *Can J Inf Libr Sci.* 2001;26.
- [44] Borsci S, Federici S, Bacci S, et al. Assessing user satisfaction in the era of user experience: comparison of the SUS, UMUX and UMUX-LITE as a function of product experience. *Int J Hum Comput Interact.* 2015;31:484–495.