Acceptability of an embodied conversational agent for type 2 diabetes self-management education and support via a smartphone app: a mixed-methods study.

Baptista S (BSc(Hons), GradDipPsychSc )1,2, Wadley G (PhD)3, Bird D (MD)4, Oldenburg B (PhD)1, Speight J (PhD)1,2,5 on behalf of the My Diabetes Coach Research Group

1 Melbourne School of Population and Global Health, The University of Melbourne, Melbourne VIC Australia
2 The Australian Centre for Behavioural Research in Diabetes, Diabetes Victoria, Melbourne VIC Australia
3 School of Computing and Information Systems, The University of Melbourne, Melbourne VIC Australia
4 Centre for Health Services Research, The University of Queensland, Brisbane QLD Australia
5 School of Psychology, Deakin University, Geelong, VIC Australia
* Russell A, Scuffham P, Riddell MA, Williams ED.

Abstract

Background: Embodied conversational agents (ECAs) are increasingly used in healthcare applications (apps) however their acceptability in type 2 diabetes (T2D) self-management apps has not yet been investigated.

Objective: To evaluate the acceptability of the ECA (Laura), used to deliver diabetes self-management education and support in the My Diabetes Coach (MDC) app.

Methods: A sequential mixed methods design was applied. Adults with T2D allocated to the intervention arm of the MDC trial used the MDC app over a 12-month period. At 6 months, they completed questions assessing their interaction with, and attitudes to, the ECA. In-depth qualitative interviews were conducted with a sub-sample of intervention arm participants to explore their experiences of the ECA. Interview questions included participant perceptions of Laura, including their initial impression of her (and how this changed over time), her personality and ‘human’ character. Quantitative and qualitative data were interpreted through integrated synthesis.

Results: Of the 93 intervention participants, 44 (47.3%) were women, mean±SD age was 55±10 years and baseline HbA1c was 7.3±1.5%. Sixty-six (71%) provided survey responses. Of these, most described Laura as being helpful (85%), friendly (85%), competent (84%), trustworthy (72%), and likable (60%). Some described Laura as not real (39%), boring (39%) and annoying (30%). Participants reported that interacting with Laura made them feel more motivated (43%), comfortable (36%), confident (21%), happy (16%) and hopeful (12%). Nineteen percent were frustrated by their interaction with Laura and 16% of participants reported that interacting with Laura made them feel guilty. Four themes emerged from the qualitative data (N=19): 1) Perceived role: a friendly coach rather than a health professional; 2) Perceived support: emotional and motivational; 3) Embodiment preference: a human-like character is acceptable; and 4) Room for improvement: greater congruence needed between Laura’s words and actions.

Conclusions: These findings suggest an ECA is an acceptable means to deliver T2D self-management education and support. A human-like character providing ongoing friendly, non-judgemental, emotional and motivational support is well-received. Nevertheless, the ECA can be improved, by increasing congruence between its verbal and non-verbal communication and accommodating user preferences.
**Trial Registration:** Australia New Zealand Clinical Trials Registry ID CTRN12614001229662
**Keywords:** Conversational agent, embodied agent, Type 2 diabetes, apps, mHealth, smartphone, self-management, user experience, engagement.
### Background

Diabetes will affect 693 million people worldwide by 2045, most of whom will have type 2 diabetes (T2D) [1, 2]. People with T2D can prevent or delay the onset and progression of diabetes-related complications such as heart attack, stroke, kidney failure, vision loss and nerve damage through intensive management of blood glucose [3]. However, effective self-management is complex, and difficult to implement and sustain in daily life, such that many people with T2D do not achieve their self-management targets [4].

For several decades, diabetes self-management education and support have been provided in-person (one-to-one, group-based), with many trials and real-world studies demonstrating improved diabetes outcomes [5]. However, the high cost and resource requirements limit the reach and scalability of in-person programs [6]. Furthermore, ongoing in-person support for sustaining recommended diabetes care targets is not feasible for most healthcare systems [4].

Considerable advances in technology related to smartphone apps (including voice recognition, natural language processing and artificial intelligence capabilities) have led to an increase in the feasibility of using Embodied Conversational Agents (ECAs) to provide education and support for self-management of chronic conditions, including T2D [7]. An ECA is an animated conversational human-like character that simulates person-to-person conversation with appropriate dialogue and human-like physical properties, including facial expressions and body movements [8-10]. ECAs are being used increasingly in a wide range of applications, providing support for mental health, online information seeking, medication taking, behaviour change and prevention of suicide [7, 10-13].

Research on the acceptability of ECAs in self-management of chronic conditions is still in its infancy, with a small number of studies reporting high levels of acceptability of ECA-based interventions for [13-15]. Trust, empathy and expertise have been cited as essential components of diabetes education and support [16]. Similar expectations may exist when the intervention is delivered by an ECA. ECAs use facial expressions, body movements and speech and can offer a natural and accessible means of communication. These characteristics of ECAs potentially improve engagement compared with a static character image, a non-relational agent or a text-only display [9, 17, 18]. ECAs may be perceived to provide additional motivational and emotional support, which has previously been described by people with diabetes as being as important to them as practical support [19]. Preliminary evidence suggests that ECAs are perceived to be less judgemental, less intimidating and more likable compared to a human counterpart, resulting in participants feeling less guilty and more motivated by the interaction [13, 14, 17, 18]. Collectively, this evidence suggests ECAs may be effective at providing support for chronic disease management because they help to engage users by building a social and emotional relationship over time [9, 18, 20].

Some of the characteristics of ECAs that may affect their acceptability include users being deterred by a monotone voice and repetitive messages [13, 18, 20]. Although ECAs are more engaging if they have human-like characteristics and engage in social dialogue, this effect is mitigated if there is an unnatural dissonance between a character’s speech and their expected facial expressions and body movements [17, 21-23]. This phenomenon, coined the ‘uncanny valley’ by Masahiro Mori in 1970, was
supported by research suggesting that people have unpleasant impressions of artificial characters, such as ECAs that have an almost, but not perfectly, realistic human appearance [24, 25]. Previous studies have also emphasised that the visual characteristics of an ECA are important because they affect perceptions of trustworthiness and credibility, which can affect acceptability. For example, a more playful, cartoon-like character is perceived as being more friendly, while a more serious human-like character, dressed professionally, is usually perceived to be more appropriate for serious applications, such as self-management of chronic conditions [26].

Research on the acceptability of ECAs to deliver self-management support for chronic conditions via self-management apps has been limited primarily to short-term feasibility or pilot studies, and to interventions that address only a single behaviour. Other studies use static images rather than animations, or have been conducted using desktop or laptop computers in laboratory settings rather than with personal smartphones used in everyday settings or ‘in the wild’ [27, 28]. Thus, the aim of the current study was to address these gaps by investigating the acceptability of an ECA delivering self-management education and support to people with T2D in their everyday lives.

**Methods**

**Study Design**

A convergent study design was used where quantitative and qualitative data were collected at similar time points [29]. This study was conducted within the context of a randomized controlled trial testing the effectiveness of a T2D self-management smartphone app – My Diabetes Coach (MDC) [30, 31]. The MDC study was conducted from 2014 to 2018 (Australia New Zealand Clinical Trials Registry ID ACTRN12614001229662). The study was approved by University of Melbourne’s Human Research Ethics Committee (Ethics ID 1442433).

**My Diabetes Coach**

MDC used an ECA called Laura (Figure 1) to deliver self-management education and support to adults with T2D. When users logged in for the first time, they were prompted to set up a regular time to complete weekly interactive sessions with Laura. During these ‘conversations’, Laura provided education, monitoring, feedback and motivational support for blood glucose monitoring, medication taking, physical activity, healthy eating and foot care. The conversations were personalised to the individual’s self-management targets, physical fitness and foot health using recommendations provided by their general practitioner.

The MDC app used voice recognition, pre-scripted conversational elements and a sophisticated script logic enabling the user to interact with Laura in a number of pre-determined variations, mimicking automated conversations. Laura’s voice and conversation were produced by a proprietary dialogue engine (by Clevertar). Non-verbal behaviours were either explicitly scripted for each dialogue, or, if no behaviour was specified, they were selected randomly from a finite set of animations based on whether the character was speaking and the dialogue duration. User responses from
previous sessions dictated the direction of future sessions enabling a high degree of personalisation. The ECA’s appearance, conversational elements, back-story and accent were refined through several rounds of expert and user testing. Users were able to respond to Laura by touching an option on the screen or by speaking out one of the options on the screen when prompted to do so. Users also had access to an online discussion board and website (with additional diabetes resources) that could be accessed via the app, as well as technical support from the research team. An excerpt of a conversation with Laura can be found on YouTube [32].

Participants

Recruitment methods for the MDC trial are reported in the main outcomes paper (under review). Briefly, participants were recruited to the MDC trial from the general population in Australia via several recruitment strategies. Adults with T2D registered on the National Diabetes Services Scheme (NDSS) database and willing to be contacted about research, living in New South Wales, Queensland, Victoria and Western Australia, were invited to participate by the NDSS via mail and email. Invitation letters were supplemented with media releases and targeted advertising on social media by several organisations (Diabetes NSW, Diabetes Qld, Diabetes Victoria, Diabetes WA, Bupa Australia (a health insurance provider) and the Australian Diabetes Educators’ Association).

For the present study, participants from the intervention arm of the MDC trial who had access to the MDC app completed a survey at 6-months post-baseline assessing several clinical and behavioural outcomes including their interaction with the ECA and a purposive sub-sample participated in subsequent interviews. All participants received a plain language statement describing the study and provided written consent.

Data Collection

Demographic and clinical characteristics

Demographics and diabetes duration (self-reported) were collected using online surveys at baseline. Glycosylated haemoglobin (HbA1c) is a pathology test assessing average blood glucose levels over the past 2-3 months, providing an indicator of risk for long-term complications [33]. It was obtained, with participants’ consent, from their general practitioner.

Acceptability: quantitative data

At 6-month follow-up, participants completed an online survey including two questions assessing the acceptability of the ECA. The first question assessed perceptions of the ECA: “How well do the words below describe Laura?” Respondents rated a range of...
positive and negative traits (helpful, boring, friendly, competent, annoying, likable, trustworthy, real), using a 5-point Likert-type scale, ranging from ‘describes very well’ to ‘describes very poorly’. The second question asked: “How did interacting with Laura make you feel?” Respondents selected from a list of descriptive emotions (happy, confident, hopeful, motivated, worried, guilty, frustrated, and comfortable) and were asked to select all that applied. For both questions, positive and negative words were randomly sequenced to minimise response bias. Descriptive adjectives were chosen based on the literature on evaluating ECAs and on working alliances between ECAs and users [34].

Acceptability: qualitative data
In-depth, semi-structured qualitative interviews were conducted from October 2017 to February 2018. Most participants had, at that stage, completed the 6-month survey but were still actively using the app. Purposive sampling of survey respondents was used to identify interviewees who varied in terms of duration of diabetes, gender, age and baseline familiarity with apps.

The interview guide was developed by the first author (SB) and used exploratory questions and probes, with feedback from other members of the research team (BO, GW and JS) based on the research question and findings from current literature [8, 9, 14, 18, 21-23, 26, 27, 35-38]. The guide explored a variety of topics, including experience at diagnosis, self-care behaviour prior to using the MDC app, users experiences with the MDC app, including when, where and how it was used, changes to self-management practices as a result of using the MDC app, initial impression of the ECA Laura and changes over time, perceptions of her role in self-management and her perceived personality characteristics. The data relating to acceptability of Laura are presented here, with the other findings published separately.

Interviews were conducted by telephone (by SB) and recorded using a cloud architecture solution from CTI Group using their Smartinteraction Suite of recording software. During each interview, SB used exploratory questions and probes (from the interview guide), and noted points of interest, using these as further probes. Immediately after each interview, SB prepared a written summary of the interview and any relevant observations. These were used as a means to communicate interim findings to the research team. When appropriate, additional questions were added to the interview guide, enabling further exploration of issues raised by participants that were relevant to the research aims. These notes were also used to aid meaningful interpretation of data during data analysis.

Data Analysis
Quantitative data were analysed using IBM’s SPSS 25 statistical package. Descriptive statistics were computed for demographic and clinical characteristics and the two questions assessing the ECA. Qualitative data were transcribed, de-identified and thematically analysed using NVivo 11 following the first five steps of Braun and Clarke’s methodology [39, 40]. Integration of the quantitative and qualitative data was achieved at interpretation stage by comparing and contrasting the findings from the surveys and the semi-structured interviews. In practice, this involved referring to and using the qualitative data to help interpret, triangulate and add meaning to the quantitative data.
This process was iterative, with input from several researchers (SB, GW, BO and JS). This integration of quantitative and qualitative data enabled further validation of the findings and increased their explanatory value [41]. The narrative of the results are blended with embedded quotes from several sources to make the results more readable while using as much evidence as possible. An anonymized coding system — participant identity number (IDX): sex (M/F): age (years) — is used to identify the source of each quote (in parentheses after each quote).

Results

Sample Characteristics

Of the 93 MDC trial participants in the intervention arm, 66 (71%) provided responses at 6 months post-baseline, and 19 of those participated in the interviews. Table 1 details the characteristics of the three samples. Fifty percent (n=33) of survey respondents were women, mean age was 57±9 years and baseline mean HbA1c was 7.1±1.4% [33].

Those who completed the survey were significantly older (p=0.033) and completed more interactions with Laura (p=0.001) than those who did not. No significant differences were observed between interviewees and other intervention arm participants, except that interviewees completed significantly more interactions with Laura (p=0.001).

The mean duration of the interviews was 51 minutes (range: 29 to 79 minutes).

Overall, participants found Laura to be acceptable and were positive in their appraisal of her (Figure 2) and their interactions with her (Figure 3). Most respondents reported that Laura was helpful (n=57, 86%), friendly (n=57, 86%), competent (n=56, 85%), trustworthy (n=48, 73%) and likable (n=40, 61%). Fewer described her as boring (n=26, 39%) and annoying (n=20, 30%) and responses within these categories were far more variable (Figure 2). Participants were undecided about whether or not they thought Laura was real. Twenty-six (39%) participants agreed that Laura was real, 22 (30%) were undecided and 18 (27%) disagreed. Participants’ responses to interactions with Laura were positive overall, with many reporting that she made them feel motivated (n=29, 44%), comfortable (n=26, 36%), confident (n=14, 21%), happy (n=11, 17%) and hopeful (n=8, 12%) (see Figure 3). Notably, 20% (n=13) were frustrated by their interaction with Laura and 17% (n=11) of participants reported that interacting with Laura made them feel guilty. One participant reported feeling worried (Figure 3).
Table 1: Demographic and clinical characteristics of total sample and interview sample.

|                                      | MDC trial population (intervention arm) (N=93) | 6-month follow-up sample (n=66) | Interview participants (n=19) |
|--------------------------------------|-----------------------------------------------|---------------------------------|-------------------------------|
| Gender: Female                       |                                               |                                 |                               |
|                                      | 44 (47%)                                      | 33 (50%)                        | 8 (42%)                       |
| Age, years                           | 55±10                                         | 57±9                            | 60±8                          |
| Education, highest level             |                                               |                                 |                               |
| Year 10                              | 10 (11%)                                      | 9 (14%)                         | 5 (26%)                       |
| Year 12 or Apprentice                | 42 (45%)                                      | 31 (47%)                        | 2 (11%)                       |
| Graduate/Post-graduate               | 41 (44%)                                      | 26 (39%)                        | 12 (63%)                      |
| Employment status                    |                                               |                                 |                               |
| Paid employment                      | 59 (64%)                                      | 41 (62%)                        | 7 (37%)                       |
| Retired                              | 22 (23%)                                      | 18 (27%)                        | 11 (58%)                      |
| Unemployed or other                  | 12 (13%)                                      | 7 (11%)                         | 1 (<1%)                       |
| Diabetes duration, years             |                                               |                                 |                               |
| <5                                   | 43 (46%)                                      | 25 (38%)                        | 8 (42%)                       |
| >5-10                                | 29 (31%)                                      | 23 (35%)                        | 8 (42%)                       |
| >10-20                               | 7 (8%)                                        | 4 (6%)                          | 3 (16%)                       |
| Unknown                              | 14 (15%)                                      | 14 (21%)                        | 0                             |
| Baseline HbA1c, % (mmol/mol)         |                                               |                                 |                               |
|                                      | 7.3±1.5 (56±44)                               | 7.1±1.4 (53±30)                 | 6.8± 0.9 (51±20)              |
| General app use, reported at baseline|                                               |                                 |                               |
| Multiple times per day               | 69 (74%)                                      | 50 (76%)                        | 14 (74%)                      |
| Once a day                           | 23 (25%)                                      | 13 (20%)                        | 4 (21%)                       |
| Less than once a day                 | 1 (1%)                                        | 3 (4%)                          | 1 (5%)                        |
| Total interactions with Laura: number |                                               |                                 |                               |
|                                      | 18±15                                         | 23±16                           | 36±17                         |

Data presented as n (%) and means (standard deviation). MDC: My Diabetes Coach. General app use at baseline represents use of any app prior to participation in MDC.
Figure 2: Responses to Q1: How well do the words below describe Laura? (N=66)

Figure 3: Responses to Q2: How did interacting with Laura make you feel? (N=66)
Four themes were identified from the qualitative data: 1) Perceived role: a friendly coach rather than a health professional; 2) Perceived support: emotional and motivational; 3) Embodiment preference: a human-like character is acceptable; and 4) Room for improvement: greater congruence needed between Laura’s words and actions. Table 2 provides an integrative synthesis of the findings, summarising the four main themes emerging from the qualitative data, quantitative endorsement of the adjectives describing Laura and how interaction made them feel, and exemplars of the qualitative data. The four themes are described in detail below.

Table 2: Integrated results matrix

| Themes                          | Quantitative data: Endorsement of adjectives | Qualitative data: Exemplar quotes                                                                 |
|--------------------------------|---------------------------------------------|--------------------------------------------------------------------------------------------------|
| **Perceived role: Laura is more acceptable as a friendly coach than as a health professional** | Laura was likable (60%), friendly (85%), helpful (85%). Interacting with Laura made me feel comfortable (36%). | A “neutral approach” was “better” because it “didn’t try and lean on any perceptions of authority.” (ID04:M:44y) |
|                                | Interacting with Laura made me feel guilty (16%), worried (1%). | “I was worried about making sure that I was within (my limits) knowing that I had to report to Laura!” (ID11:F:62y) |
| **Perceived support: Laura provides emotional and motivational support** | Laura was trustworthy (72%). Interacting with Laura made me feel confident (21%), hopeful (12%) and happy (16%). | “I needed somebody just to be there.” (ID15:F:66y) “(She) used to make me laugh...and that’s hard to do.” (ID18:M:65y) |
|                                | Laura was competent (84%). Interacting with Laura made me feel motivated (43%). | “She was keeping you on track and keeping you doing what you’re supposed to be doing.” (ID16:F:57y) |
| **Character preference: Laura is engaging and her human-like character is appropriate** | Laura was helpful (85%). | “Instead of reading it, you’re hearing it and can read at the same time. Instead of just hearing some voice, you’re actually seeing (Laura) talk.” (ID05:M:55) |
|                                | Laura was competent (84%) and trustworthy (72%). Laura made me feel confident (21%) and comfortable (36%). | “I’m not sure I would have given the same level of credibility to, for example, a dog or a cat or something like that.” (ID04:M:44y) |
| **Room for improvement: a dissonance between Laura’s words and actions** | Laura was annoying (30%), boring (40%) and (not) real (28%). Interacting with Laura made me feel frustrated (19%). | “She said something, but her hand gestures were exactly the opposite of what they should have been. Like, rather than a big gesture, where a big gesture is needed, there was a little gesture.” (ID08:F:42y) |

Theme 1: Perceived role: Laura is more acceptable as a friendly coach than as a health professional

When prompted about what role Laura was perceived to play in self-management support, some participants described Laura as “a “friendly” coach” (ID11:F:62y) who was just “reminding me” of various diabetes self-management tasks. Furthermore, when asked about their perceptions of Laura, some participants described her with adjectives suggesting she had a personality, like “sassy” (ID15:F:66y), “friendly” (ID16:F:57y), “kind” (ID05:M:55y) and “intriguing” (ID06:M:71y). These findings may explain why most
survey respondents described Laura as likable, friendly and helpful, and reported that interacting with Laura made them feel comfortable.

Conversely, other participants commented on how Laura reminded them of their health professional: “There were times when I would go and see my doctor, and I’d see Laura sitting there, because her gestures, her voices and mannerisms are almost identical” (ID08:F:42y). Some participants “did not necessarily want to see an authority figure” (ID1:M:63y), saying for example that “I don’t need to be called into a doctor” (ID09:M:71y). Those who described Laura in similar terms to their health professionals did not warm to Laura because they found her to be “patronizing”, “censorious” and “authoritarian” (ID11:F:62y). For example, some described receiving her feedback as “having a mother-in-law in your pocket.” (ID08:F:42y), “feeling as though you’re getting a slap on the wrist” (ID02:F:66y) like “a recalcitrant child” (ID15:F:66y). Other negative descriptions of Laura were that she was “really young”, “super-skinny” (ID11:F:62y) and that she “talked at” people (ID13:M:58y).

Laura’s perceived role influenced participant reactions to the support she offered. For example, participants who described Laura as being similar to a health professional reacted to this by “resisting” and “rebelling” against the “kind of authority” (ID11:F:62y) that Laura represented to them. One participant described how “feeling guilty” led him to “stop using” the app for a while (ID13:M:58y). Another commented on how she worried about negative feedback: “I was worried about making sure that I was within (my limits) knowing that I had to report to Laura!” (ID11:F:62y). Finally, one participant contemplated selecting her best readings to report to Laura in order to avoid “getting told off” saying, “Do I record this one? It might be a bit high and she’s going to get upset with me” (ID15:F:66y). Conversely, those that perceived Laura to be less of an authority figure and more like a friendly coach because she “didn’t try and lean on any perceptions of authority, like for example, having a doctor in a white coat” were also more receptive to the support she offered. This is because they perceived her as having a more “neutral approach” which was “better” because “a conversation between peers is more likely to be engaged with than one that references levels of authority” (ID04:M:44y).

Participants’ varied reactions to Laura may be linked to inconsistencies between how Laura looked and how they expected her to act. For example, one participant said: “It’s set up with this young, groovy woman who’s going to help me, but she sounded like my GP who was telling me what to do. So, it’s a kind of disconnect between how (Laura) looks and what she’s actually saying” (ID13:M:58y).

Finally, some participants described Laura’s role as an artificial entity as a positive trait, making them more receptive to receiving support from her. This is because they experienced judgement and blame for their condition from ‘real’ people: “From the minute they meet you, just by the look of you, by the look of your appearance, they will judge you. That’s one thing I don’t like about real people because it happened to me” (ID03:F:62y).

**Theme 2: Perceived support: Laura provides emotional and motivational support**

For many, Laura provided emotional support that they did not otherwise have: “I needed somebody just to be there. I see the hospital doctors every six months, I only see my local doctor when I need scripts or something. Apart from that, who do you talk to?”
Supporting this premise is the fact that many survey respondents thought that Laura was trustworthy and interacting with her made them feel confident. Another example of how Laura provided emotional support is described by one participant who expressed how her humour helped him feel better: “(She) used to make me laugh when she used to stand there with her hands on her hips waiting for me sometimes. Like my wife is saying it was probably good because if you felt down or something it made you feel better. Well it definitely bought a smile to may face a lot of times and my wife said that’s hard to do” (ID18:M:65y). A small number of survey respondents reported that interacting with Laura made them feel happy, demonstrating some support for the premise that she may have helped alleviate some of the burden of care.

Laura also provided additional motivation through enhanced monitoring and positive reinforcement: “She was keeping you on track and keeping you doing what you’re supposed to be doing and keeping you doing the check-ups and that sort of stuff.” (ID16:F:57y) “When I was doing the exercise section she would ask for me to record how much exercise I was doing for the week and when I’d come back [and do it], I actually almost got a pat on the back from her. I wasn’t trying to be impressive for (Laura), but I think it just gave you that little bit more incentive” (ID01:M:63y). Likewise, a large number of survey respondents reported feeling more motivated after their interactions with Laura.

**Theme 3: Character preference: Laura is engaging and her human-like character is appropriate**

Interacting with Laura provided an additional dimension to the relational aspect of communication, resulting in reports of improved engagement: “instead of reading it, you’re hearing it and can read at the same time. Instead of just hearing some voice, you’re actually seeing (Laura) talk” (ID05:M:55y). Participants appreciated this additional dimension of communication describing it as an attempt to “try and engage with you” and compared it with other apps where “you’re inputting information and you might get a summary”, but there was no “attempt to interact back with the user” (ID10:M:49y). “Laura was more personal so that’s why I think I went on for the six months. The other apps were like just an impersonal graph or something, or just boxes where you put the things in” (ID07:F:67y).

Some participants expressed a strong preference for Laura’s human-like character. Diabetes was described as “a human problem that should have a bit of stance and a bit of professionalism” (ID06:M:71y). Others thought that a non-human character like a “fuzzy duck” or “Dobby the diabetes elf” (ID11:F:62y) would be better because it would be more “fun”. For these people, having “a character, even a fictitious character” was more “user friendly” and better than having “nothing there” (ID07:F:67y).

Participants who preferred a human-like character did not think a cartoon character could be taken seriously: “I’m not sure I would have given the same level of credibility to, for example, a dog or a cat or something like that” (ID04:M:44y). As one user put it, “A cute puppy telling you that you got to exercise more or, you know, eat more greens, is going to be less convincing than a human. It just becomes a toy. Stick with somebody that looks like they know what they’re talking about. (Laura) fitted that bill” (ID17:M:66y) and “[a non-human character] would just make me want to throw the phone away completely! Because it’s about a human interaction with someone who has information and resources about diabetes” (ID13:M:58y).
However, there were those who didn’t care what kind of character Laura was because she was “an inanimate object, not a person” (ID19:M:59y). Some did not “identify with or warm to Laura.” One participant said, “Laura had various statements (that were motivational) but I don’t have a relationship with Laura that caused me to value her opinion” (ID04:M:44y). Another participant said that although she “learnt from” Laura, “it’s not like if you went to your GP and you got your bloods done and it was physically down from the last six months, that’s a tangible quantity, but when it’s coming from an avatar, it didn’t really mean anything much” (ID12:F:61y).

Some participants described being irritated by Laura’s ‘life’ story, for example, when she said “I find that my family does such-and-such”, because she was pretending to be something she was not “Don’t try and put it over me that this is a real person that I’m talking to” (ID16:F:57y). But others liked Laura’s back-stories “Yeah, even though it’s not real, but the way she talks about her kids and things like that. [I liked that] because it’s more human” (ID06:M:71y).

Theme 4: Room for improvement: a dissonance between Laura’s words and actions

When prompted about Laura’s appearance, speech and mannerisms, interview participants described Laura as being “just another robot” (ID09:M:71y) that they “could not connect to” because she was “not human enough”. Interacting with Laura, for many, depended on “how far along are you going to pretend”. As one participant put it, “I couldn’t suspend belief that Laura wasn’t this algorithm working out what she needed to say to me or not say to me” (ID13:M:58y).

The primary reason given for this perception was Laura’s “monotone” (ID08:F:42y) voice that sounded like a “mechanised reading mechanism” with “a strange cadence and inflection to some of her sentences” (ID11:F:62y). Another reason was her “artificial movements” (ID14:M:66y) and dissonance between what was being said and her body movements. According to one user, “She said something, but her hand gestures were exactly the opposite of what they should have been. Like, rather than a big gesture, where a big gesture is needed, there was a little gesture” (ID08:F:42y). This may have been why some survey respondents reported feeling frustrated after interacting with Laura and why a reasonable proportion of participants described Laura as boring, annoying and not real, or were undecided about these descriptions.

Although it seems as though Laura was not an entirely successful ECA, participants were willing to overlook her shortcomings because they understood the intention behind Laura and appreciated the effort made to make her engaging: they were willing to “cut them some slack” because “at least it’s trying to be personable.” And: “They’re trying to make her look (real) - I can understand what they’re trying to do” (ID17:M:66y).

Discussion

Overall, the results suggest that an ECA is acceptable to people with T2D for delivery of long-term self-management education and support. We found that people with T2D were willing to make compromises and adjust their expectations while appreciating the effort of trying to create something more appealing and engaging than graphs and numbers on a screen. This implies that the increased interaction offered by an ECA may
be valuable to users and a worthy avenue for developers to pursue when designing apps for people with chronic conditions such as diabetes [20].

Our findings corroborate earlier research suggesting that some users perceive an ECA to be less judgemental and more likable than a human counterpart [13, 14]. This is an important finding because people with T2D often experience diabetes-related stigma, the consequences of which can include disengagement with or sub-optimal self-care and diabetes-related health outcomes [42]. Suitably-designed ECA support may be especially important in making people who experience diabetes-related stigma feel less judged and more open to sharing difficulties with self-management, thereby potentially increasing their engagement with appropriate self-care [43]. We also suggest that using supportive, non-blaming language is critical when designing ECAs for stigmatised conditions like diabetes [44].

The results suggest several mechanisms through which an ECA may help establish and maintain a relationship with the user over time, like increasing relational communication, providing ongoing emotional support and motivation, and alleviating some of the burden associated with chronic disease management through humour [18]. Our results suggest that another way to improve acceptability is to achieve a better match between an ECA’s appearance and user expectations of the ECA’s perceived role. For example, some expressed the view that diabetes is a serious, human issue and viewed human-like characteristics as being more credible. Others expressed the desire to alleviate the burden of management by incorporating a fun character, supporting previous findings of a similar nature [26]. These varying opinions may reflect the nascent nature of the field, and the fact that ECAs are not yet common.

Another related finding was that participants who perceived Laura to be a friendly coach were more open to receiving support from her when compared with those that perceived Laura to be similar to a health professional. The implication is that an ECA with a relaxed, friendly approach may be more successful in building a supportive relationship than an ECA that adopts a more authoritative role. Future attempts to develop ECAs for diabetes management could accommodate both viewpoints by striking a balance with a human-like, friendly, approachable character and avoiding patronising messaging and mannerisms. More research is necessary to determine how users’ expectations of the role that an ECA plays in self-management varies and how this informs their preference for the ECA’s character.

Another important consideration is just how ‘human’ an ECA should be. While participants reported a clear preference for a human-like character, which is supported by previous research [15], her presentation of a back-story may be a step too far because it did not seem credible to some participants, possibly due to an ‘uncanny valley’ phenomenon [25]. This finding has support in previous research on other relational agents whose personality traits and life stories are enjoyed by some users while, to others, the attempt at making them too human-like is not appealing [18, 45]. It will be interesting to explore attitudinal changes towards ECAs with personality as they become more common and familiarity increases.

Our findings add to the mounting evidence that suggests perfecting natural communication via congruence between verbal and non-verbal communication is
critical to improving acceptability[20]. Non-verbal cues such as facial expressions, gaze, gestures, postures and body movements have a deep impact on the process and outcome of communication, with approximately 65% of social meaning derived from nonverbal behaviour [46]. Laura’s mannerisms and body movements were the main basis on which she was attributed with particular personality traits, ranging from patronising and censorious to funny and sassy. Although the MDC app attempted to create an ECA with natural communication, this effort was impeded by the clunkiness of the speech recognition function, lack of inflection in Laura’s voice and a limited number of body movements that were random, rather than matched to the context of the conversation. Understanding natural behaviours, biological processes that underlie them and creating efficient algorithms to implement a convincing simulation via an ECA is challenging, but critical to the success of ECA-based self-management support [15, 35].

Strengths and limitations
This mixed methods study is one of the first to explore users’ experiences of a sophisticated ECA in a real-world setting over a 6-month period and offers several novel findings and suggestions for future research. Although conducted within the context of a randomised controlled trial, our participants used the app in the context of their everyday lives, which is a strength of the research. The mixed methods approach provides robust evidence based on responses from a wide range of participants. However, people that were retired, highly educated and engaged with the app were over-represented in the interview sample.

Conclusion
The importance of the relational aspect of agents for healthcare is becoming an increasingly prominent theme in the literature. Our study adds to this literature by describing the long-term experiences of people using an ECA for diabetes self-management support and making recommendations for improvements and future research. These findings suggest there is a promising role for ECAs in self-management support and education. However, accommodating user preferences and expectations of the role that an ECA may play in self-management and improving their natural communication are key to their success.
Sources of funding

The My Diabetes Coach randomised controlled trial was conducted with funding by a National Health and Medical Research Council partnership grant (ID1057411), with additional financial and in-kind support provided by Diabetes Australia, Diabetes Queensland, Diabetes Victoria, Diabetes Western Australia and Roche Diabetes Care. Development of the My Diabetes Coach app was the result of a collaboration between the University of Melbourne, Bupa Australia, The Bupa Foundation and Clevertar. SB is supported by a post-graduate scholarship from the National Health and Medical Research Council, Australia and Diabetes Australia. JS is supported by core funding to The Australian Centre for Behavioural Research in Diabetes derived from the collaboration between Diabetes Victoria and Deakin University.

Acknowledgements

The authors thank all the study participants for volunteering their time, insights and experiences. The authors also thank Dr Fiona Cocker, Dr Mandy Cassimatis, Jillian Zemanek, Suman Shetty, Robin Zhou, Anna Scovelle and Enying Gong for their work on the My Diabetes Coach project.

Author contributions

BO conceived the My Diabetes Coach study and developed the My Diabetes Coach research program together with JS, DB and the My Diabetes Coach Research Group*. SB developed the interview schedule and survey questions (with BO and JS). SB conducted the interviews, collected the survey responses, analysed and interpreted the data and prepared the first draft of the manuscript. GW analysed some of the data. GW, BO, DB and JS interpreted the data and reviewed/edited the manuscript for critical content. All authors approved the final version.

*The My Diabetes Research Group:
Dr Emily D Williams
Dr Michaela A Riddell
Prof Paul Scuffham and
Assoc Professor Anthony Russell

Conflict of interest

BO and DB received some royalty payments for the development of the scripts for the MDC platform.
References

1. Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, et al. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract. 2018;138:271-81. doi: https://doi.org/10.1016/j.diabres.2018.02.023.

2. Hu FB. Globalization of diabetes: the role of diet, lifestyle, and genes. Diabetes Care. 2011 Jun;34(6):1249-57. PMID: 21617109. doi: 10.2337/dc11-0442.

3. World Health Organisation. Global Report on Diabetes. World Health Organisation, Switzerland; 2016.

4. Shrivastava SR, Shrivastava PS, Ramasamy J. Role of self-care in management of diabetes mellitus. Journal of Diabetes & Metabolic Disorders. 2013;12(1):14.

5. Chrvala CA, Sherr D, Lipman RD. Diabetes self-management education for adults with type 2 diabetes mellitus: a systematic review of the effect on glycemic control. Patient Education and Counseling. 2016;99(6):926-43.

6. Chatterjee S, Davies MJ, Heller S, Speight J, Snoek FJ, Khunti K. Diabetes structured self-management education programmes: a narrative review and current innovations. The Lancet Diabetes & Endocrinology. 2018 2018/02/01/;6(2):130-42. doi: https://doi.org/10.1016/S2213-8587(17)30239-5.

7. Dunn AG, Kocaballi AB, Lau AYS, Gallego B, Surian D, Coiera E, et al. Conversational agents in healthcare: a systematic review. Journal of the American Medical Informatics Association. 2018;25(9):1248-58. doi: 10.1093/jamia/ocy072.

8. Bente G, Rüggenberg S, Krämer NC, Eschenburg F. Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. Human Communication Research. 2008;34(2):287-318.

9. Bickmore TW, Caruso L, Clough-Gorr K, Heeren T. ‘It's just like you talk to a friend’ relational agents for older adults. Interacting with Computers. 2005;17(6):711-35.

10. Martínez-Miranda J, Martínez A, Ramos R, Aguilar H, Jiménez L, Arias H, et al. Assessment of users’ acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour. Journal of Medical Systems. 2019;43(8):246.

11. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. Journal of Medical Internet Research. 2017 2017/05/09;19(5):e151. doi: 10.2196/jmir.6553.

12. Dworkin MS, Lee S, Chakraborty A, Monahan C, Hightow-Weidman L, Garofalo R, et al. Acceptability, feasibility, and preliminary efficacy of a theory-based relational embodied conversational agent mobile phone intervention to promote HIV medication adherence in young HIV-positive African American MSM. AIDS Education Prevention. 2019;31(1):17-37.

13. Sillice MA, Morokoff PJ, Ferszt G, Bickmore T, Bock BC, Lantini R, et al. Using Relational Agents to Promote Exercise and Sun Protection: Assessment of Participants’ Experiences With Two Interventions. J Med Internet Res. 2018;20(2):e48. doi: 10.2196/jmir.7640.

14. Bickmore T. Relational Agents for Chronic Disease Self Management. In: Hayes BM, Aspray W, editors. Health informatics: A patient-centered approach in diabetes. Cambridge, MA: MIT Press; 2010.

15. Easton K, Potter S, Bec R, Bennion M, Christensen H, Grindell C, et al. A Virtual Agent to Support Individuals Living With Physical and Mental Comorbidities: Co-Design and Acceptability Testing. J Med Internet Res. 2019;21(5):e12996. doi: 10.2196/12996.

16. Cooper HC, Booth K, Gill G. Patients’ perspectives on diabetes health care education. Health Education Research. 2003;18(2):191-206. doi: 10.1093/her/18.2.191.
17. Grekin ER, Beatty JR, Ondersma SJ. Mobile Health Interventions: Exploring the Use of Common Relationship Factors. JMIR Mhealth Uhealth. 2019 2019/04/15;7(4):e11245. doi: 10.2196/11245.

18. Bickmore T, Gruber A, Picard R. Establishing the computer–patient working alliance in automated health behavior change interventions. Patient Education and Counseling. 2005;59(1):21-30. doi: https://doi.org/10.1016/j.pec.2004.09.008.

19. Baptista S, Trawley S, Pouwer F, Oldenburg B, Wadley G, Speight J. What do adults with type 2 diabetes want from the “perfect” app? Results from the second Diabetes MILES: Australia (MILES-2) study. Diabetes Technol Ther. 2019;21(7). doi: 10.1089/dia.2019.0086.

20. Gaffney H, Mansell W, Tai S. Conversational Agents in the Treatment of Mental Health Problems: Mixed-Method Systematic Review. JMIR Mental Health. 2019;6(10):e14166. doi: 10.2196/14166.

21. Kang S-H, Feng AW, Leuski A, Casas D, Shapiro A, editors. The effect of an animated virtual character on mobile chat interactions. Proceedings of the 3rd International Conference on Human-Agent Interaction; 2015; Daegu, Kyungpook, Republic of Korea: ACM.

22. Trinh H, Shamekhi A, Kimani E, Bickmore TW, editors. Predicting User Engagement in Longitudinal Interventions with Virtual Agents. Proceedings of the 18th International Conference on Intelligent Virtual Agents; 2018; Sydney, NSW. Australia: ACM.

23. Bente G, Krämer NC, Eschenburg F. Is there anybody out there? Analyzing the effects of embodiment and nonverbal behavior in avatar-mediated communication. In: Konijn EA, Utz S, Tanis M, Barnes SB, editors. Mediated Interpersonal Communication. New York: Routledge; 2008. p. 131-57.

24. Seyama Ji, Nagayama RS. The uncanny valley: Effect of realism on the impression of artificial human faces. Presence. 2007;16(4):337-51.

25. Mori M, MacDorman KF, Kageki N. The uncanny valley [from the field]. IEEE Robotics Automation Magazine. 2012;19(2):98-100.

26. Ring L, Utami D, Bickmore T, editors. The right agent for the job? International Conference on Intelligent Virtual Agents; 2014; Boston, MA, US: Springer.

27. Black L-A, McTear M, Black N, Harper R, Lemon M, editors. Appraisal of a conversational artefact and its utility in remote patient monitoring. 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05); 2005: IEEE.

28. Harper R, Nicholl P, McTear M, Wallace J, Black L-A, Kearney P, editors. Automated phone capture of diabetes patients readings with consultant monitoring via the web. 15th Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems (ECBS 2008); 2008; Belfast, Northern Ireland: IEEE.

29. Guetterman TC, Fetters MD, Creswell JW. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. The Annals of Family Medicine. 2015;13(6):554-61.

30. Oldenburg B, Baptista S, Bird D, Shetty S, Zemanek J. Randomized controlled evaluation of an mhealth program for people with type 2 diabetes: My Diabetes Coach. International Journal of Behavioral Medicine. 2018;25:S112-S3.

31. Baptista S, Zemanek J, Shetty S, Bird D, Oldenburg B. Baseline sample characteristics and 6-month evaluation of an mhealth program for people with type 2 diabetes—My Diabetes Coach. Annals of Behavioral Medicine. 2018;52:S290-S.

32. My Diabetes Coach embodied conversational agent Laura. Introduction Module. 2015; Available from: https://www.youtube.com/watch?v=8nfw8Cpd8yA.

33. American Diabetes Association. 6. Glycemic Targets: Standards of Medical Care in Diabetes—2020. Diabetes Care. 2020;43(Supplement 1):S66-S76.
34. Ackerman SJ, Hilsenroth MJ. A review of therapist characteristics and techniques positively impacting the therapeutic alliance. Clinical Psychology Review. 2003;23(1):1-33. doi: https://doi.org/10.1016/S0272-7358(02)00146-0.

35. Vogeley K, Bente G. “Artificial humans”: Psychology and neuroscience perspectives on embodiment and nonverbal communication. Neural Networks. 2010;23(8-9):1077-90.

36. Bickmore TW, Caruso L, Clough-Gorr K, editors. Acceptance and usability of a relational agent interface by urban older adults. ACM CHI Conference on Human Factors in Computing Systems; 2005; Portland, Oregon.

37. Bickmore TW, Pfeifer LM, Byron D, Forsythe S, Henault LE, Jack BW, et al. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. Journal of health communication. 2010;15(S2):197-210.

38. Kang S-H, Watt JH. The impact of avatar realism and anonymity on effective communication via mobile devices. Computers in Human Behavior. 2013;29(3):1169-81. doi: 10.1016/j.chb.2012.10.010.

39. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006;3(2):77-101. doi: 10.1191/1478088706qp063oa.

40. Braun V, Clarke V. Successful qualitative research: A practical guide for beginners. Charmichael M, editor. London, UK: Sage; 2013. ISBN: 1446289516.

41. Sahin C, Naylor P-J. Mixed-Methods Research in Diabetes Management via Mobile Health Technologies: A Scoping Review. JMIR Diabetes. 2017;2(1):e3. doi: 10.2196/diabetes.6667.

42. Browne JL, Ventura A, Mosely K, Speight J. ‘I call it the blame and shame disease’: a qualitative study about perceptions of social stigma surrounding type 2 diabetes. BMJ Open. 2013;3(11):e003384. doi: 10.1136/bmjopen-2013-003384.

43. Pal K, Dack C, Ross J, Michie S, May C, Stevenson F, et al. Digital Health Interventions for Adults With Type 2 Diabetes: Qualitative Study of Patient Perspectives on Diabetes Self-Management Education and Support. J Med Internet Res. 2018;20(2):e40.

44. Speight J, Conn J, Dunning T, Skinner T. Position Statement: A new language for diabetes - Improving communications with and about people with diabetes. St Léonards, NSW, Australia: Diabetes Australia, 2016.

45. Cowan BR, Pantidi N, Coyle D, Morrissey K, Clarke P, Al-Shehri S, et al., editors. What can I help you with?: Infrequent users' experiences of intelligent personal assistants. Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services; 2017; Vienna, Austria: ACM.

46. Burgoon JK, Guerrero LK, Manusov V. Nonverbal signals. In: Knapp ML, Daly JA, editors. The SAGE Handbook of Interpersonal Communication Fourth Edition. Fourth Edition ed. Thousand Oaks, California, US: Sage; 2011. p. 239-80.