Reasons to Be Satisfied or Dissatisfied With a Virtual Coach for Quitting Smoking and Becoming More Physically Active: A Mixed-Methods Analysis

Omar Sheasha
Supervisor(s): Nele Albers, Dr. Willem-Paul Brinkman
EEMCS, Delft University of Technology, The Netherlands

June 17, 2022

A Dissertation Submitted to EEMCS faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering

Reasons to Be Satisfied or Dissatisfied With a Virtual Coach for Quitting Smoking and Becoming More Physically Active: A Mixed-Methods Analysis

Omar Sheasha[1]
Supervisor(s): Nele Albers[1] , **Dr. Willem-Paul Brinkman**[1]
[1]EEMCS, Delft University of Technology, The Netherlands

Abstract

A virtual coach, Sam, was created by the Perfect Fit project to aid users in establishing a foundation to stop smoking via the means of making them more physically active. Data from 500 users had been collected from them during stages which start with a pre-screening phase until a post-questionnaire phase. In the latter stage, they provided a rating regarding their satisfaction or dissatisfaction with Sam and a free-text response mentioning the reasons why.

Through different means of analysing the data, the research question is answered, which is determining the reasons to be satisfied or dissatisfied with a virtual coach for quitting smoking and becoming more physically active. The three methods of analysis that have been used are a qualitative thematic analysis of the free-text responses, a quantitative analysis of user characteristics and ratings to obtain Pearson correlations, and a literature study.

The findings display five themes which altered users' satisfaction when using Sam. These themes are technical competency, beneficiality, user experience, interrelations and conversationalist. The quantitative analysis had resulted in correlations of at most a moderate degree. However, from the ones which displayed a significant p-value, only 3 correlations against ratings were found and 3 correlations against themes were found. A literature study conducted on outcomes of research from relevant eHealth applications aided the direction of some correlations, such as ones with personality traits. Moreover, their correlation outcomes were compared to shed light on whether the outcomes of this research and previous ones align or differ. Some correlations did align such as with the case of the beneficiality theme against a user's characteristic of conscientiousness and emotional stability.

1 Introduction

A downward global trend in the percentage of smokers existed on average, however, that is quite the contrary when it comes to lower-income and lower-socioeconomic countries or ethnic groups [1], [2].

It is important to help society in tackling this outstanding issue since smoking does not only affect its user but also greatly affects the surrounding people. This is known as second-hand smoke and it is greatly detrimental to others' health [3], [4].

Cessation methods in the present day involve more software-based techniques. Most commonly, text-based ones, also known as conversational agents or virtual coaches, have been implemented such as in the UK [5] and China [6]. Such studies, in addition to [7]–[9], have proved that eHealth interventions are successful and worthwhile to an extent. This paper delves into such a coach which was developed by the Perfect Fit project, Sam [10], [11].

User satisfaction plays a role in using these apps and elements which have been identified contribute to it were how motivating messages were and the usability of the app itself [12]. Moreover, it is not only the structure of intervention and the inclusion of motivation, however, it is also the content, length and how personal these motivational messages were that play a role in satisfaction [13].

This research aims to further fill the gap in providing insight and recommendations on what users value most about in a conversational agent which made their experience satisfactory or the contrary. This is the aspect to which this research contributes through the means of a mixed-method analysis.

The main data set from which a subset of 500 users were used for this research is publicly available [14]. These 500 were only used since they provided valid data across all stages of inputting data.

Three methods of analysis have been used, namely, a qualitative thematic analysis of the free-text responses, a quantitative analysis of user characteristics and ratings to obtain Pearson correlations, and a literature study.

The findings display 5 themes which altered users' satisfaction when using Sam. These themes are technical competency, beneficiality, user experience, interrelations and conversationalist. The quantitative analysis had resulted in correlations of at most a moderate degree. However, from the ones which displayed a significant p-value, only 3 correlations against ratings were found and 3 correlations against themes were found. A literature study conducted on outcomes of research from relevant eHealth applications aided

the direction of some correlations, such as ones with personality traits. Moreover, their correlation outcomes were compared to shed light on whether the outcomes of this research and previous ones align or differ. Some correlations did align such as with the case of the beneficiality theme against a user's characteristics of conscientiousness and emotional stability. Further insight into Sam and the methods of analysis are provided in section 2.

2 Methodology

More than 500 users took part in an experiment in which they used a text-based virtual coach, also known as a conversational agent called Sam [11]. Users were not required to stop smoking throughout this experiment. Rather, its purpose was to lay down a foundation which enables them to stop smoking down the line. Sam provided physical activities for a user to carry out as well as self-reflection activities that provided motivation and purpose to stop smoking, such as envisioning its benefits or a better future self.

Users took part in five conversational sessions where the aforementioned activities were provided and feedback was obtained, such as effort spent on activities or experience from their previous session.

A variety of data had been obtained from the participants, which are publicly accessible in this data set [14]. These are categorised and referred to as follows:

1. "Pre-screening" data: obtained in the initial stages of the experiment. Some data obtained regards a participant's background information and habits, whilst other data were used to determine participation eligibility.
2. "Pre-questionnaire" data: regards a user's characteristics which are further elaborated in [11] and in section 2.2. For example, it can offer insight into users' characteristics in terms of smoking and physical activity behaviour.
3. "Post-questionnaire" data: obtained after a participant completed all five coaching sessions. They provide a rating from -5 to 5 and a free-text response to the question 'How satisfied or dissatisfied were you with the conversational agent Sam?'.

Method triangulation

To answer the research question in determining the reasons to be satisfied or dissatisfied, three methods of analysis were used, hence a mixed-method analysis.

An inherent part of how the outcomes of these analyses are presented and tied into each other is method triangulation [15]. Method triangulation entails using outcomes of conducting different methods of analysis to either re-enforce findings or shed light on unexpected results.

The current section provides insight into what the three methods of analysis were and how they were conducted. Section 3 delves into what their outcomes were and applies method triangulation.

2.1 Thematic analysis

Out of all the participants, 500 valid responses had been obtained. These valid responses are ones in which users appropriately replied to questions and passed all the attention checks.

Thematic analysis [16], a form of qualitative data analysis, had been conducted on the 500 free-text post-questionnaire responses to generate an initial set of codes. A code represents an element of motivation or rationale behind a given response. Each response can have anywhere from zero codes to multiple codes assigned to it.

These codes are based on which components could be directly extracted from a response, using the semantic approach [16] or inferred, using the latent approach [16], which add value and relate to answering the research question.

Step 1: Getting familiar with data

At the start of the analysis, it was important to have a repeated reading [17] of the free-text responses obtained. This was to ensure all responses were valid and to get an initial grasp on concepts users pointed out regarding whether they were satisfied or not. Moreover, it provided an initial idea of patterns that exist which aided in determining how to approach the first iteration of creating codes and how granular they should have been.

Step 2: Creating codes

The first iteration of generating these codes was quite broad and served the purpose of encompassing as many elements as possible. As a result, there were codes with overlapping meanings, ones with no clear distinction, as well as codes with a low frequency which did not further add value to answering the research question.

The second iteration served the purpose of refining the codes to become more differentiable and distinct. A pitfall in using the latent approach is striking the fine balance between assumptions and inferring meanings, which can be easily blurred. It was crucial for the quality of data to not assign meanings a coder views as plausible when in fact it is a meaning the user did not intend to portray. Otherwise, this introduces a bias from the coder's perspective and generates misconstrued codes.

Step 3: Second coder

To ensure that the coding scheme was replicable and that bias was minimised, a second coder had also independently assigned codes, from the same set of codes, generated beforehand, to the same set of responses.

The second coder had been trained on 3 sets of 10 responses. These responses were chosen to demonstrate a wide array of codes. The remaining responses were independently coded by the second coder with no further guidance. According to [15] this is a form of investor triangulation which adds benefits of confirming findings, or providing new insights and perspectives into the analysis conducted.

Step 4: Finalising the coding scheme

A Cohen's Kappa had been calculated to determine the agreement between both coders. The Cohen's Kappa calculated, using this tool [18], was 0.64, which indicated a moderate agreement [19].

Finally, a discussion had been held between both coders to discuss their rationale regarding the codes and what was

deemed to be most suitable to reach a consensus on the final coding scheme. Codes were reassigned accordingly to the responses so that thematic analysis can proceed to defining and naming themes.

Step 5: Defining and naming themes

Codes were compared against each other to determine underlying similarities in their overarching purposes, which are referred to as themes. The themes identified served as a representative of the main driving factors of why users could be satisfied or dissatisfied with using a virtual coach.

Accessing analysis outcomes

The thematic analyses conducted only examined 500 valid free-text responses. The quantitative analysis was only conducted on the user characteristics of the same 500 users. The outcomes of the thematic analysis along each of the steps are publicly accessible [20]. Moreover, the main data set, which is more extensive, from which these 500 user data points were obtained is publicly available [14] as well.

2.2 Quantitative data analysis

A quantitative analysis had been conducted on data gathered from users across all stages, namely pre-screening, pre-questionnaire and post-questionnaire data.

The analysis conducted aims to find Pearson correlations, using [21] as a guideline for the interpretation of the degree of correlation, between users' characteristics, ratings, and corresponding themes in their free-text responses. These user characteristics from the pre-questionnaire data often consisted of numerous components that had to be aggregated to adequately provide complete insight. Based on the guidelines by [22] and [23], the calculated Cronbach's α for each of the aggregated characteristics, were significantly high, using the means of the items as index measures.

User characteristic	**Cronbach's α**
Physical activity identity	0.89
Quitting self identity	0.76
Smoker group identity	0.79
Smoker self identity	0.74
Non-smoker self identity	0.67
Non-smoker group identity	0.79
Quitting self efficacy	0.83
Need for cognition (NFC)	0.77

Table 1: Cronbach's α for aggregated user characteristics used within this research.

2.3 Literature study

As the final step of this research, a literature study had been conducted on studies which delve into user experiences of similar eHealth applications. Findings from this literature study helped re-enforce results obtained or bring to light differences that exist. Either of the outcomes provides further insight into eHealth and what can be done to further improve such applications.

3 Findings

This section provides insight into what each of the themes represents. Some codes may be used to provide further clarity, however, the full list of codes per theme used can be found in [20]. The themes generated are as follows:

1. Technical competency
2. Beneficiality
3. User experience
4. Interrelation
5. Conversationalist

The following table displays Pearson correlations and the p-value of hypotheses tested against themes. As previously mentioned, [21] is used as a guideline in determining the degree of correlation.

Hypothesis	**Correlation**	**p-value**
Age vs Themes		
Technical competency	0.07	0.13
Beneficiality	0.08	0.07
User experience	0.11	0.01
Interrelation	0.03	0.56
Conversationalist	0.02	0.61
Conscientiousness vs Themes		
Technical competency	0.04	0.43
Beneficiality	0.03	0.47
User experience	0.14	0.001
Interrelation	0.03	0.54
Conversationalist	0.05	0.23
Emotional stability vs Themes		
Technical competency	0.004	0.93
Beneficiality	0.06	0.18
User experience	0.11	0.01
Interrelation	0.004	0.93
Conversationalist	0.06	0.16
Openness to experiences vs Themes		
Technical competency	0.008	0.86
Beneficiality	0.04	0.42
User experience	0.04	0.32
Interrelation	0.03	0.57
Conversationalist	0.06	0.21

Table 2: Pearson correlation coefficient for hypotheses tested against themes and their corresponding p-value.

This research conducted covers an age range of 18 to 74 years old. Therefore, it was interesting to explore whether age played a role in this research and displayed any significant correlations with the themes generated since it also displays a wider age range than previous research which is outdated, however. Given time constraints, no recent research was found explicitly tackling this correlation again, therefore, it was worth filling in this gap found.

On the other hand, personality traits have been shown to affect perceived usefulness and satisfaction in a user's experience [24], [25]. As a result, the remaining three hypothesis categories contain an element of a user's personality trait

against themes. This is to be able to compare results with previous research and denote points of alignment or differences. Additionally, the data set used in this research is wider, therefore, it is capable of providing further insight into these correlations.

The reasoning for these chosen traits was that they seemed to go hand in hand with the characteristics a person changing their lifestyle should contain. The following illustrates what each trait portrays, using previous research as insight [24], [25], and why it was of interest to test correlations against them.

- Conscientiousness: relates to mindfulness, and goal-directed behaviours of a person so that they achieve results whilst also being efficient. Correlations here were tested to shed light on whether how motivated a person is to commit to Sam and the activities provided in order to achieve results alters their satisfaction.
- Emotional stability: contains notions of how people deal with their stress, anxiety, or physical and emotional responses to threats and challenges. Correlations here were tested to shed light on whether a person's nature in how they view Sam's purpose, in an optimistic light or in a pessimistic threatening manner, alters their satisfaction.
- Openness to experiences: demonstrates how eager a person is to learn and delve into new experiences in an accepting manner. Correlations here were tested to shed light on whether how open-minded and welcoming a person is to new situations and changes, such as what Sam involves them in, alters their satisfaction.

Results from the previous studies [24], [25] are compared in their relevant sections when themes are discussed to demonstrate whether findings from this research align or differ from the ones conducted previously.

The correlations explored are listed in table 2, to provide an insight into their degree of correlation and their p-value. However, only three of them meet the p-value criteria of being less than or equal to 0.05. These three all relate to correlations against the theme 'User Experience', which is also stated again in its section. Since the remainder did not satisfy the p-value requirement of significance, they are only indicated in table 2 for transparency and further insight.

Referencing to user responses

In the published outcomes [20] each user had been assigned a "sequentialID" from P1 to P500. These will be referred to after directly quoting users' free-text responses when providing further insight into a theme.

Since each response may have multiple codes assigned to it, the part of the response which corresponded to why the theme, and thus the underlying code, were appropriate to be assigned to it are indicated in bold. However, if the entire quote relates solely to one theme, it will not be indicated in bold for readability's sake.

Not only were the responses chosen based on which ones contained the most frequent codes, but also so that the underlying codes assigned are not identical across responses quoted in order to portray a wider and clearer definition of a theme.

3.1 Technical competency

"*It was a really nice experience,* ***it felt like talking to a real human being.***" (P25)

"***Very interactive*** *and offered good activities*" (P106)

"*it was intuitive and* ***reasonably well structured***" (P279)

This theme regards users' critique on how well the technicalities of the coach were perceived. The most frequent aspect that users cared about in this regard was how realistic Sam was, either a human-like or bot-like experience. The runner-up relates to the interactivity of Sam which ties into how interesting it was to communicate with an agent, instead of only filling in a form.

This theme also encompasses technical aspects of the agent such as how relevant the information it provided was, its straightforwardness, restrictiveness in options to reply with and its efficiency, to name a few. This theme was most frequently occurring with a frequency of 208 and a total of 12 codes that fall under it.

3.2 Beneficiality

"***It felt he was helping me along***" (P22)

"*It was extremely user friendly and* ***I felt it was very supportive***" (P191)

"*He is straight to the point and* ***his/her advice and suggestions on how to approach the stopping of smoking is very helpful and informative, and easy to remember.***" (P343)

This theme regards well Sam was in achieving its purpose. This relates to how helpful it was, and the encouraging, educational and pragmatic aspect of it. Moreover, it relates to whether it provided good advice to the users, aside from how relevant it is to them which was mentioned in the previous theme. This theme consists of 5 codes with a total frequency of 111, making it the second-lowest frequency.

In previous studies [24], [25] it was shown that conscientiousness and emotional stability played a role in "perceived usefulness", which is similar to this theme's meaning, however no correlations of significant p-values were found in table 2 regarding this relationship.

3.3 User experience

"*It was nice and easy to complete*" (P219)

"*It felt quite engaging and easy to use*" (P297)

"*He did just fine.*" (P467)

This theme regards the experience of using the coach. It relates to whether it was quite enjoyable or that users felt quite neutral towards it, whether it was user friendly, or engaging. This theme also encompasses fun, simplicity, and smoothness perceived by the user. A total of 12 codes fall under this theme with a total frequency of 178, making it the second-highest.

A correlation of 0.11 and a p-value of 0.01 had been identified with this theme against age, indicating a weak correlation between the two. Moreover, a correlation of 0.14 against

conscientiousness with a p-value of 0.0001 indicates a weak correlation as well. Lastly, a correlation of 0.11 and a p-value of 0.01 against emotional stability had been identified which also indicates a weak correlation.

3.4 Interrelation

> "*It's almost like talking to a real person.* ***He's compassionate and very friendly****, and helpful.*" (P251)
>
> "*He was very sympathetic and always asked me how i was doing.*" (P316)
>
> "*It was an interesting and unique approach, very unexpected.* ***It added a personal touch to the quit smoking concept***" (P438)

This theme regards the relationship formed, or lack thereof, between the user and the coach. It relates to how personal of a connection the user felt, how friendly Sam was and its emotional intelligence. Emotional intelligence encompasses the notion of care, sincerity, empathy, kindness, and not imposing guilt onto the user.

This theme consists of 14 codes and a total frequency of 129 thus placing it as the third most frequent theme, even though it consists of the most codes.

3.5 Conversationalist

> "***Messages were well written and easy to understand.*** *Seemed interactive enough too.*" (P8)
>
> "***It was easy to converse with, and easy to understand.*** *The activities were largely useful.*" (P27)
>
> "*It was a pleasant experience,* ***very fluent and natural***" (P234)

This theme portrays the conversational skills the coach displayed when communicating with users. It regards how articulate Sam was, how easy they were to understand, and the ease of reaching out to Sam. Moreover, it encompasses how natural and responsive Sam was in conversing and that Sam provided neither pre-determined answers nor predictable.

A total of 7 codes fall under this theme with a total frequency of 106, making it the least frequent theme.

3.6 Correlations with satisfaction ratings obtained

Satisfaction ratings were obtained from users in the post-questionnaire where they provided a rating from -5 to 5 regarding the question "How satisfied or dissatisfied were you with the conversational agent Sam?". They will be referred to as only "Rating" in table 3 for formatting's sake.

- Ease is an averaged component of two elements, accounting for direction, which portrays how easy the user found it to do the assigned activities.
- Motivation is an averaged component of two elements, accounting for direction, which portrays how motivated the user was to do the assigned activities.
- The remaining characteristics' explanations can be found in [11].

When it comes to correlations with ratings provided by the 500 users from -5 to 5 regarding the question "How satisfied or dissatisfied were you with the conversational agent Sam?", only three were of a significant p-value. Those were ratings against ease, motivation and conscientiousness.

Hypothesis	**Correlation**	**p-value**
Rating - Ease	0.35	< 0.001
Rating - Motivation	0.45	< 0.001
Rating - Quitting self identity	0.003	0.95
Rating - Extraversion	0.04	0.42
Rating - Agreeableness	0.07	0.10
Rating - Conscientiousness	0.10	0.02
Rating - Emotional stability	0.02	0.71
Rating - Openness to experiences	0.01	0.81
Rating - TTM physical activity	0.04	0.43

Table 3: Pearson's correlation coefficient for hypotheses tested with ratings and their corresponding p-value.

4 Responsible Research

To further increase transparency and capability of reproducing the outcomes in this research, the initial codes assigned by the first coder, the codes assigned by the second coder, as well as the final codes and themes assigned are publicly accessible through [20]. Moreover, the main data set, from which this research only looked at a subset, is publicly accessible [14].

All hypothesis and their corresponding results have been reported regardless if they were of a satisfactory p-value for instance since they all provide valuable insight. This is to remain transparent and honest throughout the research which had been conducted.

Two types of triangulation based on [15] had been used in this research, method and researcher triangulation.

Firstly, method triangulation entails conducting different methods of analysis or studies. Results obtained from each are compared to either re-enforce findings or shed light on unexpected results. This method is applied within section 3 as an inherent part of it. This refers to how the outcomes of the three stages in methodology came together when presenting findings to form a coherent whole and interrelating results.

Secondly, researcher triangulation is a form of investor triangulation [15]. It adds the benefits of confirming findings or providing new insights or perspectives into the analysis conducted. An application of this is in section 2.1. This method of triangulation further reduces bias from each of the coders involved to result in a higher quality of codes generated, meaning they are feasible to replicate by others wishing to conduct this study. From the discussion which had been held with the second coder to finalise the codes, codes have been removed or merged on the basis of too similar meaning with no added value in their distinction.

5 Recommendations

Based upon the most prevalent codes and themes, the following aspects are suggested to be incorporated into other eHealth applications. In addition, insight is used from further more free-text responses than the ones that had been used in earlier sections as quotes.

Users value how realistic and interactive the experience was. This lays down a foundation upon which elements of

personal connection and engagement can be achieved which played a role in making users remain active with their tasks and desire to go back to them.

A personal connection goes hand in hand with the emotional intelligence displayed by the agent and its friendliness. This further enhances a user's experience by making them feel more comfortable, less judged, having a sense of familiarity established and being paid attention to. The latter also relates to providing relevant information in the user's context.

Last but not least, the agent itself should fulfil its objective of being pragmatic, helpful and providing good advice. Moreover, the experience should also be enjoyable and user-friendly.

6 Conclusions and Future Work

Three methods of analysis have been used, namely, a qualitative thematic analysis of the free-text responses, a quantitative analysis of user characteristics and ratings to obtain Pearson correlations, and a literature study.

The outcomes of thematic analysis across all stages are also publicly available [20]. The main data set from which a subset of 500 users are used for this research is publicly available [14]. These 500 were only used since they provided valid data across all stages of inputting data.

The findings display 5 themes which altered users' satisfaction when using Sam. These themes are technical competency, beneficiality, user experience, interrelations and conversationalist. The quantitative analysis had resulted in correlations of at most a moderate degree. However, from the ones which displayed a significant p-value, only three correlations against ratings were found and three correlations against themes were found. A literature study conducted on outcomes of research from relevant eHealth applications aided the direction of some correlations, such as ones with personality traits. Moreover, their correlation outcomes were compared to shed light on whether the outcomes of this research and previous ones align or differ.

Given the time constraints, a greater extent of the literature study was not feasible to be carried out for other eHealth applications. It would be very interesting to further explore what applications exist with the same purpose as Sam and compare their outcomes. These outcomes could either reenforce what this research has found or differences may be noted, which are equally valuable to determine. Some correlations did align such as with the case of the beneficiality theme against a user's characteristic of conscientiousness and emotional stability. However, in the limited time that had remained for a literature study, no extra research had been found addressing aspects of other themes found or correlations against age. This could be conducted on a larger scale in future research to come to more conclusive results of how other studies align with this research.

Moreover, the quantitative data analysis did not include many results which were of a significant p-value. This could be due to the sample size not being large enough for a correlation to be prominent. In addition, more coders could be incorporated when it comes to thematic analysis to have a more concrete code scheme especially given a larger sample size where more bias can be introduced.

When it comes to limitations, it was not plausible for any causal relationships to be drawn from the analyses conducted. This was outside the scope of this research and requires a greater level of detail using different methods to deduce them. This research is only capable of providing insight into correlations which were significant given the data set size of these 500 users. In addition, it provided insight into what aspects users value which can not be generalised, however, to any other users who were not a part of this experiment.

7 Acknowledgement

I would like to extend my appreciation to my supervisors for providing valuable feedback throughout the entirety of this research. They have always been punctual, responsive, and meticulous.

Secondly, my research peer group were such a delight to work with. We established a fantastic support system and held multiple fruitful meetings together. We were keen on each other's success and always lent out a helping hand to each other.

Last but not least, a special heartfelt thank you to my entire family for being my strongest support system since the start of my bachelor's. This journey would not have been possible without their unconditional love, emotional support, and invaluable wisdom enabling me to thrive and keep my spirits high. For that, I end this journey by thanking them and expressing my deepest gratitude to them all.

References

[1] W. H. Organization *et al.*, *WHO global report on trends in prevalence of tobacco smoking 2000-2025*. World Health Organization, 2018.

[2] F. Islami, M. Stoklosa, J. Drope, and A. Jemal, "Global and regional patterns of tobacco smoking and tobacco control policies," *European Urology Focus*, vol. 1, no. 1, pp. 3–16, 2015, ISSN: 2405-4569. DOI: https://doi.org/10.1016/j.euf.2014.10.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405456915000024.

[3] A. Besaratinia and G. P. Pfeifer, "Second-hand smoke and human lung cancer," *The lancet oncology*, vol. 9, no. 7, pp. 657–666, 2008.

[4] H. Yousuf, M. Hofstra, J. Tijssen, *et al.*, "Estimated worldwide mortality attributed to secondhand tobacco smoke exposure, 1990-2016," *JAMA network open*, vol. 3, no. 3, e201177–e201177, 2020.

[5] C. Free, R. Knight, S. Robertson, *et al.*, "Smoking cessation support delivered via mobile phone text messaging (txt2stop): A single-blind, randomised trial," *The Lancet*, vol. 378, no. 9785, pp. 49–55, 2011, ISSN: 0140-6736. DOI: https://doi.org/10.1016/S0140-6736(11)60701-0. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140673611607010.

[6] Y. Liao, Q. Wu, B. C. Kelly, *et al.*, "Effectiveness of a text-messaging-based smoking cessation intervention ("happy quit") for smoking cessation in china: A randomized controlled trial," *PLoS medicine*, vol. 15, no. 12, e1002713, 2018.

[7] S. Muellmann, S. Forberger, T. Möllers, E. Bröring, H. Zeeb, and C. R. Pischke, "Effectiveness of ehealth interventions for the promotion of physical activity in older adults: A systematic review," *Preventive Medicine*, vol. 108, pp. 93–110, 2018, ISSN: 0091-7435. DOI: https://doi.org/10.1016/j.ypmed.2017.12.026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0091743517305145.

[8] J. Grolleman, B. v. Dijk, A. Nijholt, and A. v. Emst, "Break the habit! designing an e-therapy intervention using a virtual coach in aid of smoking cessation," in *International Conference on Persuasive Technology*, Springer, 2006, pp. 133–141.

[9] M. Z. Hurmuz, S. M. Jansen-Kosterink, T. Beinema, K. Fischer, H. Op den Akker, and H. J. Hermens, "Evaluation of a virtual coaching system ehealth intervention: A mixed methods observational cohort study in the netherlands," *Internet interventions*, vol. 27, p. 100 501, 2022.

[10] N. Albers, M. Neerincx, and W.-P. Brinkman, "Reinforcement learning-based persuasion by a conversational agent for behavior change," in *33rd Benelux Conference on Artificial Intelligence and30th Belgian-Dutch Conference on Machine Learning*, 2021, pp. 729–732.

[11] N. Albers and W.-P. Brinkman, *Perfect fit - experiment to gather data for and test a reinforcement learning-approach for motivating people*, May 2021. DOI: 10.17605/OSF.IO/K2UAC. [Online]. Available: osf.io/k2uac.

[12] F. Luna-Perejon, S. Malwade, C. Styliadis, *et al.*, "Evaluation of user satisfaction and usability of a mobile app for smoking cessation," *Computer methods and programs in biomedicine*, vol. 182, p. 105 042, 2019.

[13] U. Müssener, M. Bendtsen, J. McCambridge, and P. Bendtsen, "User satisfaction with the structure and content of the nexit intervention, a text messaging-based smoking cessation programme," *BMC Public Health*, vol. 16, no. 1, pp. 1–8, 2016.

[14] N. Albers, M. A. Neerincx, and W.-P. Brinkman, "Acceptance of a Virtual Coach for Quitting Smoking and Becoming Physically Active: Dataset," May 2022. DOI: 10.4121/19934783.v1. [Online]. Available: https://data.4tu.nl/articles/dataset/Acceptance_of_a_Virtual_Coach_for_Quitting_Smoking_and_Becoming_Physically_Active_Dataset/19934783.

[15] R. Nancy Carter, D. Bryant-Lukosius, and R. Alba DiCenso, "The use of triangulation in qualitative research," in *Oncology nursing forum*, Oncology Nursing Society, vol. 41, 2014, p. 545.

[16] M. E. Kiger and L. Varpio, "Thematic analysis of qualitative data: Amee guide no. 131," *Medical teacher*, vol. 42, no. 8, pp. 846–854, 2020.

[17] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006. DOI: 10.1191/1478088706qp063oa. eprint: https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa.

[18] *Quantify interrater agreement with kappa*. [Online]. Available: https://www.graphpad.com/quickcalcs/kappa1/.

[19] M. L. McHugh, "Interrater reliability: The kappa statistic," *Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.

[20] O. Sheasha, *PerfectFit-project/analysis_acceptance_virtual_coach_satisfaction: thematic_outcomes*, version 1.3, Jun. 2022. DOI: 10.5281/zenodo.6658218. [Online]. Available: https://doi.org/10.5281/zenodo.6658218.

[21] C. P. Dancey and J. Reidy, *Statistics without maths for psychology*. Pearson education, 2007.

[22] M. Tavakol and R. Dennick, "Making sense of cronbach's alpha," *International journal of medical education*, vol. 2, p. 53, 2011.

[23] K. S. Taber, "The use of cronbach's alpha when developing and reporting research instruments in science education," *Research in science education*, vol. 48, no. 6, pp. 1273–1296, 2018.

[24] M. Alshawmar, B. Tulu, and A. H. Phillips, "Influence of personality traits on the continued use of fitness apps.," in *HICSS*, 2022, pp. 1–10.

[25] S. Devaraj, R. F. Easley, and J. M. Crant, "Research note—how does personality matter? relating the five-factor model to technology acceptance and use," *Information systems research*, vol. 19, no. 1, pp. 93–105, 2008.