Beyond calories: evaluating how tailored communication reduces emotional load in diet-coaching

Simone Balloccu and Ehud Reiter
University of Aberdeen / United Kingdom
simone.balloccu@abdn.ac.uk
e.reiter@abdn.ac.uk

Abstract

Dieting is a behaviour change task that is difficult for many people to conduct successfully. This is due to many factors, including stress and cost. Mobile applications offer an alternative to traditional coaching. However, previous work on apps evaluation only focused on dietary outcomes, ignoring users' emotional state despite its influence on eating habits. In this work, we introduce a novel evaluation of the effects that tailored communication can have on the emotional load of dieting. We implement this by augmenting a traditional diet-app with affective NLG, text-tailoring and persuasive communication techniques. We then run a short 2-weeks experiment and check dietary outcomes, user feedback of produced text and, most importantly, its impact on emotional state, through PANAS questionnaire. Results show that tailored communication significantly improved users' emotional state, compared to an app-only control group.

1 Introduction

An unhealthy diet poses a serious threat to an individual's health. Research showed that a poor diet kills more people than smoking (Afshin et al., 2019) and that obesity has tripled since 1975.[1] Coaching through human experts is one of the most effective ways to improve diet (Gordon et al., 2017; Schmittdiel et al., 2017), but it can be too expensive for disadvantaged groups, adding to other costs associated with a healthy diet (Aggarwal et al., 2011; Barosh et al., 2014; Morris et al., 2014; Håkansson, 2015).

E-health apps are a cheaper alternative, although there is mixed evidence about their effectiveness (Wang et al., 2016; McCarroll et al., 2017; Lee et al., 2018; Aromatario et al., 2019). Compared to experts, apps often show sub-optimal communication. Typically apps focus on data presentation (e.g.: charts), limiting the use of text to short and fixed messages. This could be the reason why previous apps evaluation focused primarily on diet outcomes. There has been little work on effective communication for dieting tools: this should be addressed as it plays a big role in engagement and adherence (Lee and Cho, 2017). Dieting habits are also known to be influenced by emotional state (Macht and Simons, 2011; Koenders and van Strien, 2011; Klump et al., 2016), yet no prior work on diet-apps investigated communication's role in this.

In this paper, we implement an advanced communication strategy and investigate its effect on emotional state in the context of diet coaching apps. We exploit affective-NLG, text-tailoring and persuasive communication techniques to create weekly diet reports. Reports are implemented as an additional layer on top of a standard diet app, augmenting its communicative capability. We then proceed to evaluate our system in a short experiment. We compare participants that used the report-augmented app, with an app-only control group. Unlike previous work, we do not focus our human evaluation on dietary outcome only. We inspect communication adequacy through user feedback on a variety of measures including readability and accuracy. As a novel contribution we evaluate if our reports improved participant's affective state. We adopt a validated psychometric tool, the PANAS questionnaire (Watson et al., 1988), to analyse the behaviour of both groups on a weekly basis. Participants who received our report experienced significantly more positive emotions and fewer negative ones. We also observe the opposite behaviour in the control group.

In Section 2 we expose the common limits of diet apps under the functional, communication and psychological aspect. We also briefly describe SOTA

[1] https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

Proceedings of the 2nd Workshop on Human Evaluation of NLP Systems (HumEval), pages 42 - 53
May 27, 2022

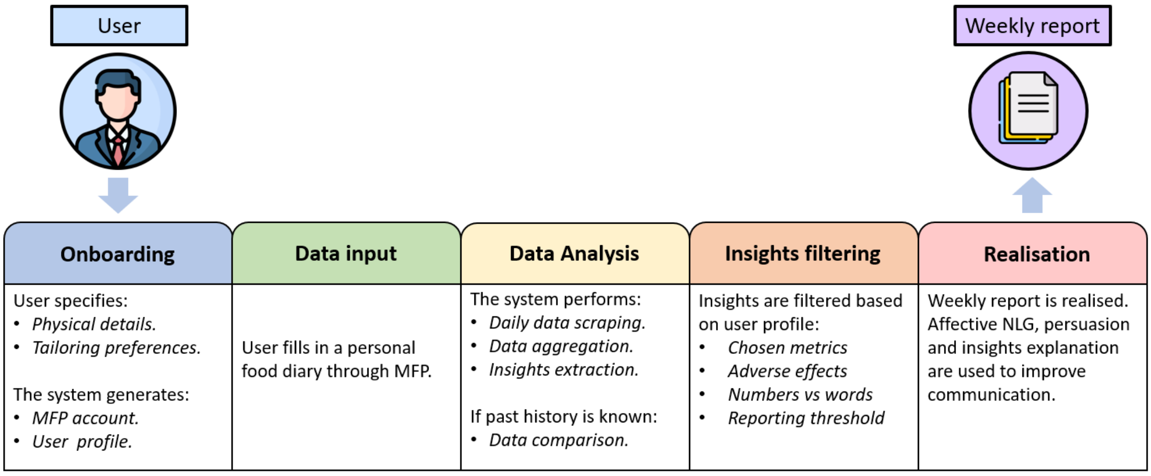


Figure 1: Execution flow, from user subscription to report delivery.

Hello Dan18777, you told us that you want to gain some weight, so we wrote this report especially for you.

Your calorie intake could use some improvement: there was an occasional lack of food (generally you ate about a third less than your target). It was a bit better the previous time, we're sure you can do it again!
Friday looks like the most problematic day (you ate about half of your target).

It seems that sodium and protein intake needs a bit of improvement.

Your sodium consumption was about half of your target. Of the foods you ate, "Spinacina" was the highest in sodium. It would be better to correct this as sodium deficiency can lead to cramps.

Also, your protein consumption was about half of your target. Last week it was better and we know you can do it again! "Pizza" was the food you ate which had the most protein. Keep in mind that protein deficiency can be responsible for muscle loss.

Figure 2: Example of a generated weekly report on the second week.

in communication-based systems for diet-coaching. In Section 3 we detail our approach to augmenting diet-coaching apps communication, and describe the implemented features. We present our experiment methodology in Section 4, and discuss the results in Section 5. In Section 6 we sum up our conclusions and present our future research directions. Finally, in Section 7 we detail the procedure through which we ensured ethical compliance for our experiment.

2 Related work

Today people can access lots of diet tools, but both academic and commercial products show some common limits. Some of these are purely functional: missing features that negatively impact on effectiveness. This includes low accuracy (Vasiloglou et al., 2020), fixed suggestions (Lieffers et al., 2018) or the excessive use of humans in the loop (Teeriniemi et al., 2018)[2,3]. Low accuracy is an obvious limit to the app's effectiveness; fixed suggestions overlook customisation and potential dangers (like allergies, user taste and religious food dogmas); major use of human experts nullifies apps' usefulness in the first place. However, these problems can be solved by expanding the tool-set of features and evaluating dietary outcomes.

But if we consider the behavioural component of dieting, we raise different problems, for example at communication and psychological level. Previous research showed that behaviour change benefits more from advanced communication (Van Dorsten and Lindley, 2008; Balloccu et al., 2021; Whitehead and Parkin, 2022) than from factual text. Diet apps (Corcoran, 2014; Evans, 2017; Tredrea et al., 2017), however, do not follow this logic and favour data presentation, through visual features (like charts, color codes and tables) (Eikey, 2021). At communication level, used text is typically short, fixed and lacks informativeness (Vasiloglou et al., 2020).

[2] www.rise.us
[3] https://www.noom.com/

A first way to improve the communication of diet apps could be the use of fine-tuned, domain-specific NLG, combined with text-tailoring (Kreuter and Wray, 2003; Noar et al., 2007) and persuasive communication (Guerini et al., 2011; Duerr and Gloor, 2021; Shabir et al., 2022). This is motivated by the relationship between personalisation and engagement in diet apps (Lieffers et al., 2018; Zmora and Elinav, 2021), and the role of persuasion in behaviour change (Orji and Moffatt, 2018; Balloccu et al., 2021). Additionally, NLG has been used in various healthcare domains (Reiter et al., 2003; Finley et al., 2018; Pauws et al., 2019; Hommes et al., 2019), including some work in nutrition. Shed (Lim-Cheng et al., 2014) is a tailored diet-system that exploits NLG to propose alternative meal plans in real time. Initial inspection of user acceptance showed it as a promising system for further evaluation. A conceptual diet-recommender system has been proposed (Ritschel et al., 2019), focusing on reinforcement learning for linguistic personalisation. Other work (Donadello et al., 2019) presented a NLG-based persuasive reasoner to address dietary guidelines violations. Evaluation showed the appropriateness of presented feedback, and its effectiveness in reducing the amount of violations compared to canned text. MADiMan (Anselma et al., 2018; Anselma and Mazzei, 2020) is a persuasive diet-coaching system, developed to convince the user to opt for an healthier diet. Evaluation in both controlled and uncontrolled scenario revealed that users appreciated the presence of both visualisations and text, and confirmed its persuasiveness. While these works evaluated the use of persuasion and dietary outcomes, we note that tailoring involved only data analysis (e.g.: custom meal plans) and not textual features. Moreover, previous research did not inspect whether the adopted communication techniques had an effect on users' emotional state. This aligns with previous evidence that diet-apps rarely consider this element (Ferrara et al., 2019). We know from nutrition research (Torres and Nowson, 2007; Puddephatt et al., 2020; Riffer et al., 2019) that user's emotional/affective state influences eating habits, causing various issues including calorie excess (Fong et al., 2019), emotional (Macht and Simons, 2011; Van Strien et al., 2012) and binge (Klump et al., 2016) eating. The importance of this factor is also confirmed by previous research of the matter in other domains such as Cognitive Behaviour Therapy (Fitzpatrick et al., 2017), mental well-being (Ly et al., 2017), substance abuse (Prochaska et al., 2021) or emotional support in public speaking (Murali et al., 2021). To the best of our knowledge, this is the first work in NLG for nutrition that investigates the influence of the system on affective state.

3 Augmenting diet apps communication

We implement an NLG report generator for diet-coaching based on our previous work (Balloccu et al., 2020a)[4], and use it to augment the communication strategy of a traditional diet app. We use MyFitnessPal (MFP) (Evans, 2017) as data source. The execution flow can be seen in Figure 1. The report is tailored based on various preferences. Users were asked to specify:

1. A nickname
2. Their motivation for using the system (e.g.: "I want to lose weight")
3. How they wanted to display quantities in reports. The options were pure values (e.g.: "50% of your calorie goal") or fuzzy quantification (e.g.: "half of your calorie goal")
4. Metrics of interest (one or more from: calories, carbohydrates, protein, fat, sodium and sugar)
5. Threshold for intake reporting. This allows the system to ignore small anomalies like 1% calorie excess.
6. Whether or not to see possible adverse effects of their dietary choices (e.g.: consequences of calorie excess/deficit)

Username and motivation are injected in the report, to make it feel more personal, while the other elements are used for content selection and tailoring. Reports are further enriched with the following insights:

1. **Worst day:** the day whose caloric intake was the furthest from the goal.
2. **Nutrients ranking:** nutrients are ranked and only the two furthest ones from the goal are shown.

[4]Code available at: `https://bitbucket.org/uccollab/diet-tailoring/`

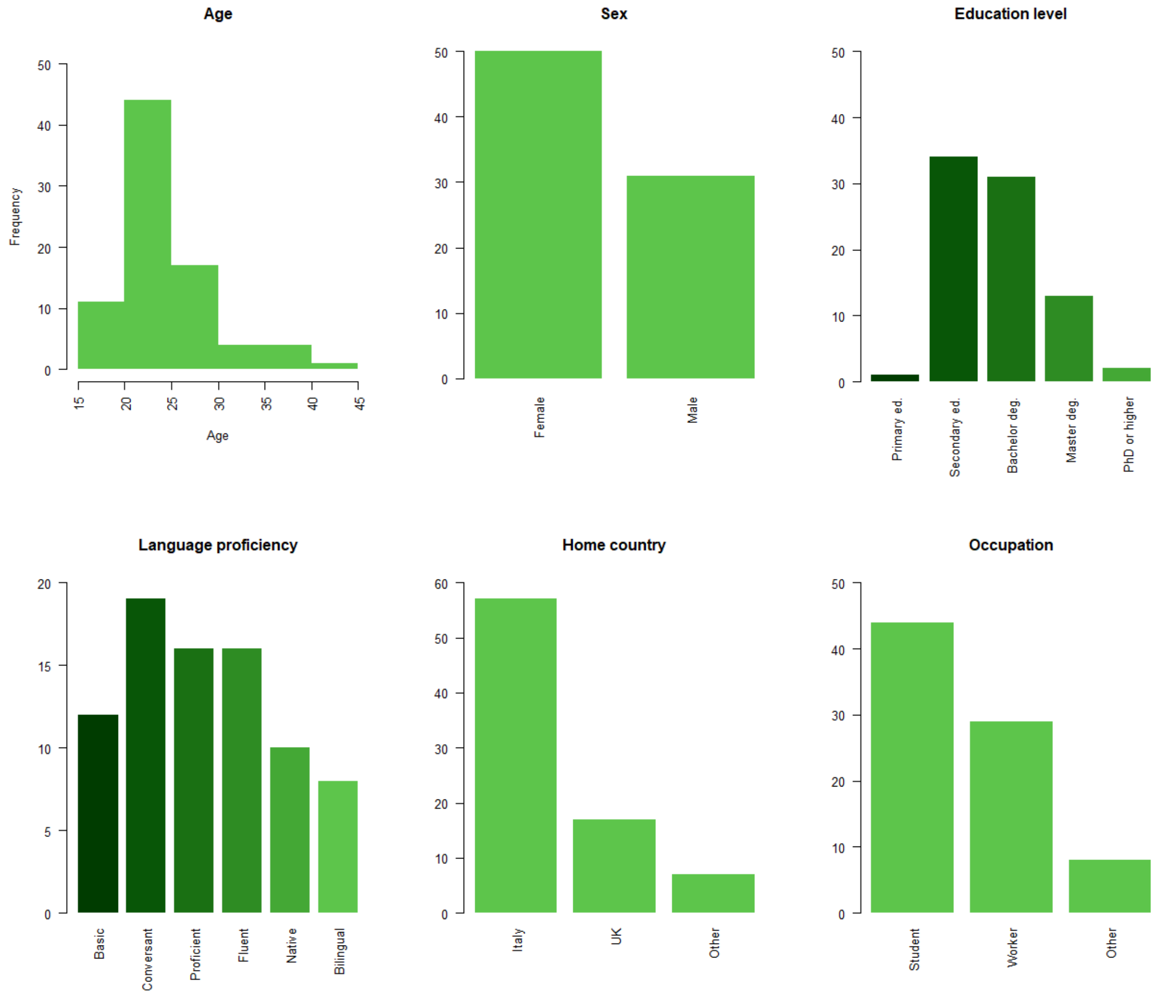


Figure 3: Population demographics. For language proficiency we adopt the scale proposed at `https://csb.uncw.edu/`. The process was supervised in order to avoid erroneous self-assessments.

3. **Food analysis:** for each nutrients, the food which provided most of it is listed.
4. **Comparisons:** if previous week data are present, intakes are compared and the eventual improvement/worsening is shown.

Finally, we adopt Affective NLG (de Rosis and Grasso, 2000; Mahamood and Reiter, 2011; Piwek, 2002), framing the document as positive-toned. This includes expressing comfort in case of negative developments and congratulations for positive ones (e.g.: calorie intake improved/worsened). Each report referred to the past week. An output example can be seen in Figure 2.

4 Experiment setup

We evaluated the effect of our reports on the diet and emotional state of users in a 2-weeks experiment. A total of 81 participants were recruited (see Section 7 for details). Population demographics can be seen in Figure 3.

Participants were trained in using MFP and asked to log their meals through the app for the following 2 weeks. They were then randomly split into two groups: "Report group" ($n = 43$) and "Control group" ($n = 38$). Participants in report group received one report at the end of each week, while control group could only see the insights provided in MFP.

About 60% of the participants (from both groups) agreed to fill-in a weekly PANAS questionnaire (Watson et al., 1988) that we used to monitor their emotional/affective state. PANAS consists of 20 mixed positive (e..g: "Attentive", "Proud", "Strong" etc...) and negative (e.g.: "Hostile", "Guilty", "Scared" etc...) words. Users score

Positive and Negative Affect Schedule (PANAS-SF)

Indicate the extent you have felt this way over the past week.		Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
PANAS 1	Interested	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 2	Distressed	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 3	Excited	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 4	Upset	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 5	Strong	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 6	Guilty	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 7	Scared	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 8	Hostile	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 9	Enthusiastic	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 10	Proud	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 11	Irritable	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 12	Alert	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 13	Ashamed	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 14	Inspired	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 15	Nervous	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 16	Determined	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 17	Attentive	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 18	Jittery	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 19	Active	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 20	Afraid	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Figure 4: Weekly PANAS questionnaire, as it was administered during the experiment

	Participants that improved (%)		
Goal	Report group	Control group	p-value (χ^2)
Calories	42%	23%	≈ 0.23
Nutrient 1	56%	33%	≈ 0.16
Nutrient 2	40%	42%	≈ 0.43

Table 1: Diet outcomes per group (after two weeks). For each group, we report how many participants got closer to their dietary goals.

	Improvement (distance from goal)		
Goal	Report group	Control Group	p-value (t-test)
Calories	+1.78%	+6.53%	≈ 0.14
Nutrient 1	-25.92%	-29.60%	≈ 0.17
Nutrient 2	-10.74%	-17.36%	≈ 0.76

Table 2: Diet outcomes per group (after two weeks): we report participants average improvement in terms of distance from dietary goals (for calories and the nutrients that were mentioned in the report). For distance from goal, a decrease is considered and improvement.

each word on a 5-points scale, based on what extent they felt that way during the past week. An example of the questionnaire can be seen in Figure 4. PANAS generates a pair of independent scores: Positive Affect (PA) and Negative Affect (NA). Each score refers to what degree the participant experienced positive (for PA) or negative (for NA) emotions. PANAS improvement is expressed as an increase in PA, a decrease in NA or both. Participants were given PANAS before the experiment and at the end of each week, and always before report delivery to avoid any influence. We note this implies that, at the end of the first week, neither the report or control group had seen a report when filling out the form. We chose PANAS as a measuring tool because its scores are generalised across multiple aspects of the affective state. Both scores include the cumulative contribution a wide range of emotions. Other tools such as SPSS (Cohen et al., 1994) or HAM-A (Hamilton, 1959) would have been too focused on specific aspects. We also avoided combining multiple tools as this could have been too tiring for participants, leading to inaccurate results. Finally, at the end of the experiment, participants were asked to evaluate the report by scoring eight Likert-7 questions that can be seen in Figure 6. Participants were also given the chance to express an open comment about the system. We let participants from the control group read one single report at the end of the experiment to let them express their feedback as well.

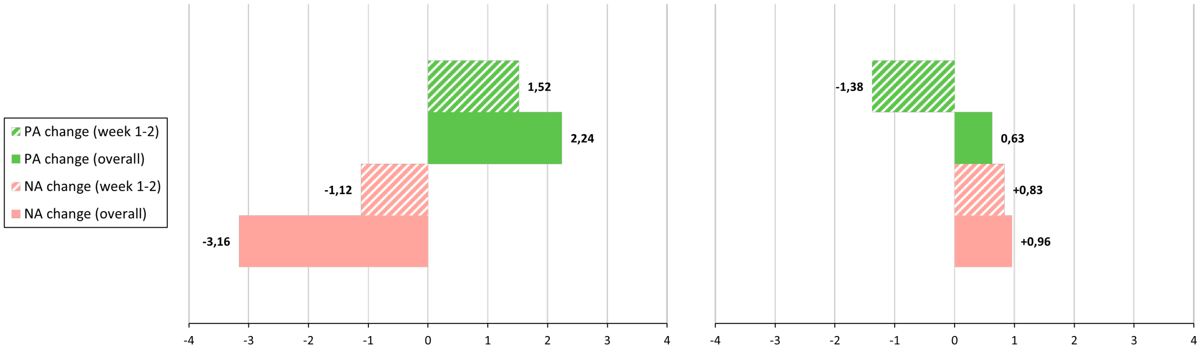
Through this setup, we inspected the following research hypotheses:

Hypothesis 1 (H1): *Participants in report group improved their diet (in terms of caloric and nutritional intakes) more than control group.*

Hypothesis 2 (H2): *Participants in report group improved their positive affect score more than control group.*

Hypothesis 3 (H3): *Participants in report group improved their negative affect score more than control group.*

While H1 is comparable to classic diet-coaching evaluation, we introduce H2 and H3 as a novel investigation of the communicative potential of these tools, related to users' emotional state. For H1 we check the initial distance between MFP goals (for calories and nutrients) and user intake. Then, we verify if, at the end of week 1 and week 2, participants got closer to said goals. For nutrients, we consider the two most unbalanced ones (those that could be seen in the report). For H2 and H3 we monitor weekly PANAS scores (PA and NA) for each group. Since no group had access to reports when completing PANAS at the end of the first week, we use this value as a starting point. Then, we check differences at the end of week 2 and overall (from the start of the experiment).



(a) **Report Group** (b) **Control group**

Figure 5: Results from PANAS analysis for both groups. We report Positive Affect (**PA**) and Negative Affect (**NA**) change from week 1 to week 2, and overall (from the start of the experiment). For PA higher is better; for NA lower is better.

5 Results and discussion

In terms of dietary outcomes, we obtained mixed results, but none of these were significant. In fact, the majority of participants improving calories and the first most unbalanced nutrient were in report group, but a chi-squared test revealed no significance (see Table 1). Both groups worsened their calories intake and we saw the biggest improvement in control group for nutrients (see Table 2). Again, a t-test revealed that none of these results is statistically significant. People in the report group were more likely to improve, while people in control group showed the biggest improvements, but a longer experiment is needed to assess whether reports (or their absence) played a role in this. With these results, we reject H1. However, it is safe to assume that reports didn't worsen the effectiveness of MyFitnessPal.

On the other hand, PANAS analysis gave us more interesting results. Initially, we verified through a t-test that the two groups shared similar initial PA (average difference = 0.1, $p = 0.96$) and NA (average difference = 1.7, $p = 0.51$). Then, we checked how scores changed for both groups. PA and NA were checked at week 1, week 2 and across the whole experiment. The report group showed bigger improvements, both in terms of PA and NA (see Figure 5).

The report group showed (through t-test and Sidak's p-value adjustment) a significantly bigger improvement for PA on the second week ($p = 0.04$) and for NA across the whole experiment ($p = 0.04$). Generally, the report group improved both scores more than the control group in any other situation, but only in these two cases the p-values were statistically significant. These results tell us that the report group tended to experience significantly:

1. More positive feelings during the second week
2. Fewer negative feelings across the whole experiment

than the control group. It is interesting to see PA significantly improving during second week. Since PANAS was administered before each report delivery, that was the first time that the report group could express their emotional state after reading a report.

The control group generally showed worse behaviour: PA greatly worsened during second week, while there was a slight improvement across the whole experiment (but much lower than the one experienced from the report group). NA consistently worsened in both cases. This tells us that the control group experienced a heavier emotional load during the experiment. We hypothesise that this is related to the cognitive load: the control group had to figure out how to interpret MFP charts and numerical data, while the report group was helped by the explanation provided in the generated text. Moreover, nutrients ranking helped participants from the report to focus on a limited amount of elements. In contrast, participants from the control group had to pay attention to calories and each nutrient. We also checked whether we could find some differences in the emotional state during the first week, when no group had access to the report. We observed a bigger PA improvement in control

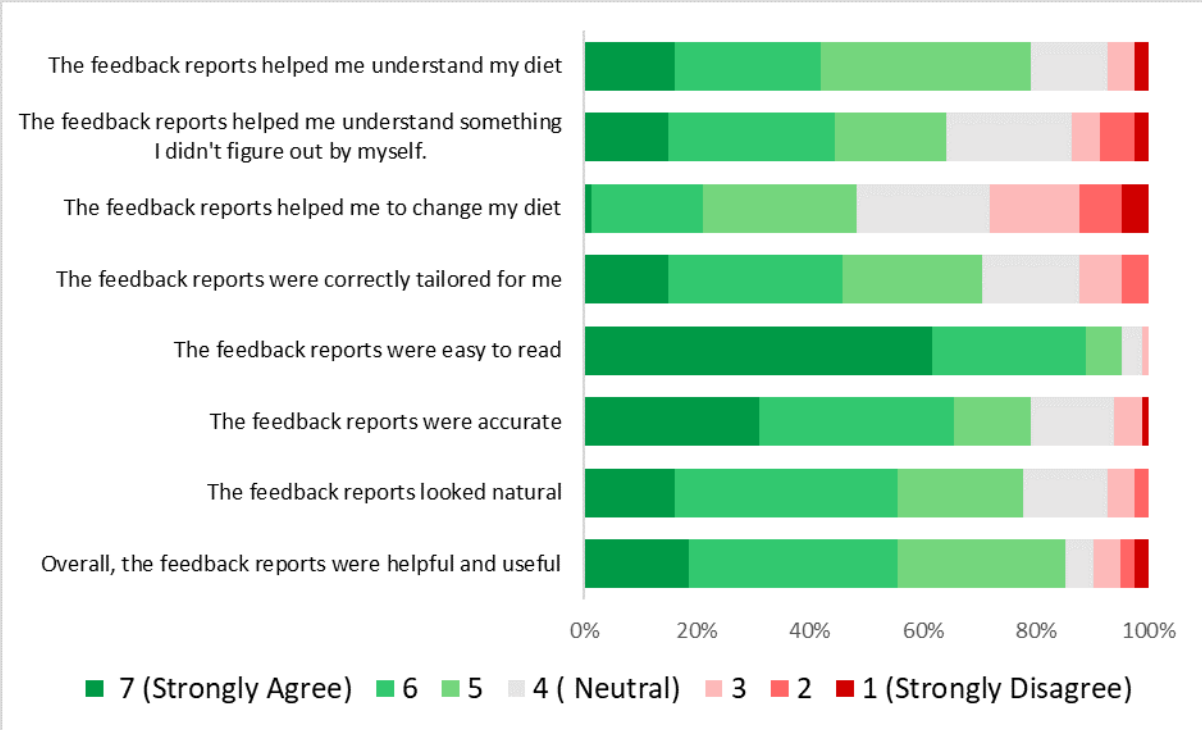


Figure 6: Overview of feedback from participants.

group ($\Delta = 2$) than in report group ($\Delta = 0.72$). The opposite happened for NA, with the report group improving it ($\Delta = -2.04$) and control group slightly worsening it ($\Delta = 0.13$). None of these was statistically significant. Considering the lack of reports, in this case, we can assume that the cognitive load was similar. Overall, we could see a significant improvement in emotional state for the report group (PA in the second week, NA overall). With these results we confirm H2 and H3.

Final feedback (Figure 6) was mostly positive. The lowest scores belong to the help in changing diet, which could also be related to the experiment duration. When given the chance to express a comment on the system, many participants asked for charts and graphical elements which could have improved understanding. This result aligns with previous research (Law et al., 2005; Molina et al., 2011; Gkatzia et al., 2017), suggesting that a combination of visual features and textual communication could be the most effective approach.

6 Conclusion

In this paper we evaluated the effects of augmented communication in diet-apps using Affective NLG, tailoring and persuasive communication techniques. Unlike previous work in evaluating diet-coaching systems, we did not look only at dietary outcomes. Since diet is influenced by psychological factors we introduced a novel evaluation by adopting a validated psychometric tool (PANAS). We inspected whether our reports could play a role in improving users' affective state.

Our hypotheses were confirmed, as we found that participants who read the report experienced more positive emotions and fewer negative ones. We also saw the opposite in most cases for the control group. Our work has shown that improved communication can reduce the impact of emotional load on dieting. Most importantly, we showed how important it is to consider the psychological component when designing, developing and evaluating communication systems, in diet-coaching and other domains. We could not see an effect on diet itself, which encourages us to run a longer trial (one month or more) in future, to further assess the effectiveness of our communication strategy. However, we ran just a basic assessment on the psychological side. We plan to expand our evaluation procedure by combining multiple tools and scales. As our previous work pointed out, stress is one particular factor that could be worth monitoring (Balloccu et al., 2020b), so this is one of the main directions we intend to follow. We also could not run any kind of ablation test. This leaves us with the conclusion that our approach did work, but without any insights on how different elements (affective NLG, persuasion or text tailoring) contributed.

Based on the feedback from users, more than just text is required to improve the system. We leave this as future work. Still feedback was largely positive with regards to textual features and comprehension. We note that the questions were not accompanied by rigorous definitions of "readability", "accuracy" and others. Users expressed feedback based on their own personal idea of these concepts and this raises questions regarding the reliability of the results. We consider the overall uniformity of ratings as an indicator that all participants had a "common" definition of the proposed concepts. Still this uncertainty contributes to a well-known problem in human evaluation (Howcroft et al., 2020), so we commit to more rigorous and uniform metric definitions in future.

7 Ethical considerations

This section sums up the procedure we adopted to ensure the ethical compliance of our experiment.

7.1 Preliminary review

Before starting the experiment, procedure and materials were carefully reviewed by the University of Aberdeen Ethics Board. Our experiment proposal was accepted without major revisions.

7.2 Recruitment

Participants were recruited through physical interaction on campus (by flyer distribution), department mailing list or social media public posts. No recruitment qualification was specified, beside the lack of health conditions that are known to affect individuals diet. This includes pregnancy, suffering from eating disorders or psychological treatments. This was done since our system has been developed to work in "standard" situations, while the aforementioned cases would have pose high risks for participants. Participants were showed a consent form containing all the information regarding the experiment procedure. All participants had to confirm their acceptance of these conditions (through check-boxes and signature) in order to proceed with the experiment. Participants were given an email contact in case of problems during the experiment.

7.3 Pay and workload

Each participants received £20 (or 20€ for participants outside of UK) at the end of the experiment, as a token of gratitude for their contribution. Access to the token of gratitude was bound to the compliance of the following condition:

1. To complete the experiment (that is, using MFP for two weeks; giving the final feedback)
2. To provide, to the best of their capabilities, the most complete and accurate food diaries they could.

Requirement 1) also included PANAS forms for those participants who agreed to do so. For 2) participants were supervised and given support about meal logging and eventual missing entries. Participants were also informed of the possibility of abandoning the experiment (up to the point of data analysis), which would result in exclusion from receiving the token of gratitude.

7.4 Data protection and storage

A MFP account for each participant was generated through temporary email that was in no way linked to their identity. Following the experiment conclusion, all accounts have been blocked. Data have been safely stored and anonymised.

References

Ashkan Afshin, Patrick John Sur, Kairsten A Fay, Leslie Cornaby, Giannina Ferrara, Joseph S Salama, Erin C Mullany, Kalkidan Hassen Abate, Cristiana Abbafati, Zegeye Abebe, et al. 2019. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet*, 393(10184):1958–1972.

Anju Aggarwal, Pablo Monsivais, Andrea J Cook, and Adam Drewnowski. 2011. Does diet cost mediate the relation between socioeconomic position and diet quality? *European journal of clinical nutrition*, 65(9):1059–1066.

Luca Anselma, Simone Donetti, Alessandro Mazzei, and Andrea Pirone. 2018. CheckYourMeal!: diet management with NLG. In *Proceedings of the Workshop on Intelligent Interactive Systems and Language Generation (2IS&NLG)*, pages 45–47, Tilburg, the Netherlands. Association for Computational Linguistics.

Luca Anselma and Alessandro Mazzei. 2020. Building a persuasive virtual dietitian. In *Informatics*, volume 7, page 27. Multidisciplinary Digital Publishing Institute.

O Aromatario, A Van Hoye, A Vuillemin, A-M Foucaut, C Crozet, J Pommier, and L Cambon. 2019. How do mobile health applications support behaviour changes? a scoping review of mobile health applications relating to physical activity and eating behaviours. *Public health*, 175:8–18.

Simone Balloccu, Steffen Pauws, and Ehud Reiter. 2020a. A nlg framework for user tailoring and profiling in healthcare. In *SmartPhil@ IUI*, pages 13–32.

Simone Balloccu, Ehud Reiter, Matteo G Collu, Federico Sanna, Manuela Sanguinetti, and Maurizio Atzori. 2021. Unaddressed challenges in persuasive dieting chatbots. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, pages 392–395.

Simone Balloccu, Ehud Reiter, Alexandra Johnstone, and Claire Fyfe. 2020b. How are you? introducing stress-based text tailoring. In *Proceedings of the Workshop on Intelligent Information Processing and Natural Language Generation*, pages 62–70, Santiago de Compostela, Spain. Association for Computational Lingustics.

Laurel Barosh, Sharon Friel, Katrin Engelhardt, and Lilian Chan. 2014. The cost of a healthy and sustainable diet–who can afford it? *Australian and New Zealand journal of public health*, 38(1):7–12.

Sheldon Cohen, Tom Kamarck, Robin Mermelstein, et al. 1994. Perceived stress scale. *Measuring stress: A guide for health and social scientists*, 10(2):1–2.

Kathleen Corcoran. 2014. Fooducate.

Fiorella de Rosis and Floriana Grasso. 2000. *Affective Natural Language Generation*, pages 204–218. Springer Berlin Heidelberg, Berlin, Heidelberg.

Ivan Donadello, Mauro Dragoni, and Claudio Eccher. 2019. Persuasive explanation of reasoning inferences on dietary data. In *PROFILES/SEMEX@ ISWC*.

Sebastian Duerr and Peter A Gloor. 2021. Persuasive natural language generation–a literature review. *arXiv preprint arXiv:2101.05786*.

Elizabeth V Eikey. 2021. Effects of diet and fitness apps on eating disorder behaviours: qualitative study. *BJPsych Open*, 7(5).

Daniel Evans. 2017. Myfitnesspal. *British Journal of Sports Medicine*, 51(14):1101–1102.

Giannina Ferrara, Jenna Kim, Shuhao Lin, Jenna Hua, and Edmund Seto. 2019. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. *JMIR mHealth and uHealth*, 7(5):e9232.

Gregory Finley, Erik Edwards, Amanda Robinson, Michael Brenndoerfer, Najmeh Sadoughi, James Fone, Nico Axtmann, Mark Miller, and David Suendermann-Oeft. 2018. An automated medical scribe for documenting clinical encounters. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 11–15, New Orleans, Louisiana. Association for Computational Linguistics.

Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial. *JMIR mental health*, 4(2):e7785.

Mackenzie Fong, Ang Li, Andrew J Hill, Michelle Cunich, Michael R Skilton, Claire D Madigan, and Ian D Caterson. 2019. Mood and appetite: Their relationship with discretionary and total daily energy intake. *Physiology & behavior*, 207:122–131.

Dimitra Gkatzia, Oliver Lemon, and Verena Rieser. 2017. Data-to-text generation improves decision-making under uncertainty. *IEEE Computational Intelligence Magazine*, 12(3):10–17.

Neil F. Gordon, Richard D. Salmon, Brenda S. Wright, George C. Faircloth, Kevin S. Reid, and Terri L. Gordon. 2017. Clinical effectiveness of lifestyle health coaching: Case study of an evidence-based program. *American Journal of Lifestyle Medicine*, 11(2):153–166.

Marco Guerini, Oliviero Stock, Massimo Zancanaro, Daniel J O'Keefe, Irene Mazzotta, Fiorella de Rosis, Isabella Poggi, Meiyii Y Lim, and Ruth Aylett. 2011. Approaches to verbal persuasion in intelligent user interfaces. In *Emotion-Oriented Systems*, pages 559–584. Springer.

Andreas Håkansson. 2015. Has it become increasingly expensive to follow a nutritious diet? insights from a new price index for nutritious diets in sweden 1980–2012. *Food & Nutrition Research*, 59(1):26932.

M Hamilton. 1959. Hamilton anxiety scale. *Group*, 1(4):10–1037.

Saar Hommes, Chris van der Lee, Felix Clouth, Jeroen Vermunt, Xander Verbeek, and Emiel Krahmer. 2019. A personalized data-to-text support tool for cancer patients. In *Proceedings of the 12th International Conference on Natural Language Generation*, pages 443–452, Tokyo, Japan. Association for Computational Linguistics.

David M. Howcroft, Anya Belz, Miruna-Adriana Clinciu, Dimitra Gkatzia, Sadid A. Hasan, Saad Mahamood, Simon Mille, Emiel van Miltenburg, Sashank Santhanam, and Verena Rieser. 2020. Twenty years of confusion in human evaluation: NLG needs evaluation sheets and standardised definitions. In *Proceedings of the 13th International Conference on Natural Language Generation*, pages 169–182, Dublin, Ireland. Association for Computational Linguistics.

Kelly L Klump, Shannon M O'Connor, Britny A Hildebrandt, Pamela K Keel, Michael Neale, Cheryl L Sisk, Steven Boker, and S Alexandra Burt. 2016.

Differential effects of estrogen and progesterone on genetic and environmental risk for emotional eating in women. *Clinical Psychological Science*, 4(5):895–908.

Paul G Koenders and Tatjana van Strien. 2011. Emotional eating, rather than lifestyle behavior, drives weight gain in a prospective study in 1562 employees. *Journal of Occupational and Environmental Medicine*, 53(11):1287–1293.

Matthew W Kreuter and Ricardo J Wray. 2003. Tailored and targeted health communication: strategies for enhancing information relevance. *American journal of health behavior*, 27(1):S227–S232.

Anna S Law, Yvonne Freer, Jim Hunter, Robert H Logie, Neil McIntosh, and John Quinn. 2005. A comparison of graphical and textual presentations of time series data to support medical decision making in the neonatal intensive care unit. *Journal of clinical monitoring and computing*, 19(3):183–194.

H Erin Lee and Jaehee Cho. 2017. What motivates users to continue using diet and fitness apps? application of the uses and gratifications approach. *Health communication*, 32(12):1445–1453.

Mikyung Lee, Hyeonkyeong Lee, Youlim Kim, Junghee Kim, Mikyeong Cho, Jaeun Jang, and Hyoeun Jang. 2018. Mobile app-based health promotion programs: a systematic review of the literature. *International journal of environmental research and public health*, 15(12):2838.

Jessica R. L. Lieffers, José F. Arocha, Kelly Anne Grindrod, and Rhona M. Hanning. 2018. Experiences and perceptions of adults accessing publicly available nutrition behavior-change mobile apps for weight management. *Journal of the Academy of Nutrition and Dietetics*, 118 2:229–239.e3.

Nathalie Rose Lim-Cheng, Gabriel Isidro G Fabia, Marco Emil G Quebral, and Miguelito T Yu. 2014. Shed: An online diet counselling system. In *DLSU research congress*, pages 1–7.

Kien Hoa Ly, Ann-Marie Ly, and Gerhard Andersson. 2017. A fully automated conversational agent for promoting mental well-being: A pilot rct using mixed methods. *Internet interventions*, 10:39–46.

Michael Macht and Gwenda Simons. 2011. https://doi.org/10.1016/j.appet.2011.10.005. In *Emotion regulation and well-being*, pages 281–295. Springer.

Saad Mahamood and Ehud Reiter. 2011. Generating affective natural language for parents of neonatal infants. In *Proceedings of the 13th European Workshop on Natural Language Generation*, pages 12–21.

Rebecca McCarroll, Helen Eyles, and Cliona Ni Mhurchu. 2017. Effectiveness of mobile health (mhealth) interventions for promoting healthy eating in adults: A systematic review. *Preventive Medicine*, 105:156–168.

Martin Molina, Amanda Stent, and Enrique Parodi. 2011. Generating automated news to explain the meaning of sensor data. In *International Symposium on Intelligent Data Analysis*, pages 282–293. Springer.

Michelle A Morris, Claire Hulme, Graham P Clarke, Kimberley L Edwards, and Janet E Cade. 2014. What is the cost of a healthy diet? using diet data from the uk women's cohort study. *Journal of Epidemiology & Community Health*, 68(11):1043–1049.

Prasanth Murali, Ha Trinh, Lazlo Ring, and Timothy Bickmore. 2021. A friendly face in the crowd: Reducing public speaking anxiety with an emotional support agent in the audience. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*, pages 156–163.

Seth M Noar, Christina N Benac, and Melissa S Harris. 2007. Does tailoring matter? meta-analytic review of tailored print health behavior change interventions. *Psychological bulletin*, 133(4):673.

Rita Orji and Karyn Moffatt. 2018. Persuasive technology for health and wellness: State-of-the-art and emerging trends. *Health informatics journal*, 24(1):66–91.

Steffen Pauws, Albert Gatt, Emiel Krahmer, and Ehud Reiter. 2019. Making effective use of healthcare data using data-to-text technology. In *Data Science for Healthcare*, pages 119–145. Springer.

Paul Piwek. 2002. An annotated bibliography of affective natural language generation. *Information Technology Research Institute (ITRI), University of Brighton, ITRI-02-02*.

Judith J Prochaska, Erin A Vogel, Amy Chieng, Matthew Kendra, Michael Baiocchi, Sarah Pajarito, and Athena Robinson. 2021. A therapeutic relational agent for reducing problematic substance use (woebot): development and usability study. *Journal of Medical Internet Research*, 23(3):e24850.

Jo-Anne Puddephatt, Gregory S Keenan, Amy Fielden, Danielle L Reaves, Jason CG Halford, and Charlotte A Hardman. 2020. 'eating to survive': A qualitative analysis of factors influencing food choice and eating behaviour in a food-insecure population. *Appetite*, 147:104547.

Ehud Reiter, Roma Robertson, and Liesl M Osman. 2003. Lessons from a failure: Generating tailored smoking cessation letters. *Artificial Intelligence*, 144(1-2):41–58.

Friedrich Riffer, Manuel Sprung, Hannah Münch, Elmar Kaiser, Lore Streibl, Kathrin Heneis, and Alexandra Kautzky-Willer. 2019. Relationship between psychological stress and metabolism in morbidly obese individuals. *Wiener klinische Wochenschrift*, pages 1–11.

Hannes Ritschel, Kathrin Janowski, Andreas Seiderer, and Elisabeth André. 2019. Towards a robotic dietitian with adaptive linguistic style.

Julie A. Schmittdiel, Sara R. Adams, Nancy Goler, Rashel S. Sanna, Mindy Boccio, David J. Bellamy, Susan D. Brown, Romain S. Neugebauer, and Assiamira Ferrara. 2017. The impact of telephonic wellness coaching on weight loss: A "natural experiments for translation in diabetes (next-d)" study. *Obesity*, 25(2):352–356.

Habiba Shabir, Matthew D'Costa, Zain Mohiaddin, Zaeem Moti, Hamza Rashid, Daria Sadowska, Benyamin Alam, and Benita Cox. 2022. The barriers and facilitators to the use of lifestyle apps: A systematic review of qualitative studies. *European journal of investigation in health, psychology and education*, 12(2):144–165.

A-M Teeriniemi, T Salonurmi, T Jokelainen, H Vähänikkilä, T Alahäivälä, P Karppinen, H Enwald, M-L Huotari, J Laitinen, H Oinas-Kukkonen, et al. 2018. A randomized clinical trial of the effectiveness of a web-based health behaviour change support system and group lifestyle counselling on body weight loss in overweight and obese subjects: 2-year outcomes. *Journal of internal medicine*, 284(5):534–545.

Susan J Torres and Caryl A Nowson. 2007. Relationship between stress, eating behavior, and obesity. *Nutrition*, 23(11-12):887–894.

Matthew S Tredrea, Vincent J Dalbo, and Aaron T Scanlan. 2017. Lifesum: easy and effective dietary and activity monitoring. *British Journal of Sports Medicine*, 51(13):1042–1043.

Brent Van Dorsten and Emily M Lindley. 2008. Cognitive and behavioral approaches in the treatment of obesity. *Endocrinology and metabolism clinics of North America*, 37(4):905–922.

Tatjana Van Strien, C Peter Herman, Doeschka J Anschutz, Rutger CME Engels, and Carolina de Weerth. 2012. Moderation of distress-induced eating by emotional eating scores. *Appetite*, 58(1):277–284.

Maria F Vasiloglou, Stergios Christodoulidis, Emilie Reber, Thomai Stathopoulou, Ya Lu, Zeno Stanga, and Stavroula Mougiakakou. 2020. What healthcare professionals think of "nutrition & diet" apps: An international survey. *Nutrients*, 12(8):2214.

Qing Wang, Bjørg Egelandsdal, Gro V Amdam, Valerie L Almli, and Marije Oostindjer. 2016. Diet and physical activity apps: perceived effectiveness by app users. *JMIR mHealth and uHealth*, 4(2):e33.

David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, 54(6):1063.

Kirsten Whitehead and Tracey Parkin. 2022. Uk dietitians' views on communication skills for behaviour change: A 10 year follow-up survey. *Journal of Human Nutrition and Dietetics*, 35(1):112–123.

Niv Zmora and Eran Elinav. 2021. Harnessing smartphones to personalize nutrition in a time of global pandemic. *Nutrients*, 13(2):422.