

Generating Motivation for Obesity Prevention

Reinforcing behaviour change with GPT-4 generated motivational text messages.

Joris Raaphorst
Industrial Design
Technical University Eindhoven
Eindhoven
j.raaphorst@student.tue.nl

ABSTRACT

In healthcare behaviour change is a common goal, to prevent the emergence or worsening of chronic diseases and early mortality (Davidson & Scholz, 2020). It is evident that unhealthy behaviours such as poor diet and lacking physical exercise can lead to overweight and obesity, and even result in non-communicable chronic diseases, such as diabetes and cardio-vascular diseases (Farhud, 2015).

In addition, a growing global shortage of healthcare professionals (WHO, 2019) dealing with aging societies, calls for actions to enable and support health-related behaviours change on individual levels.

With the emergence of conversational Artificial Intelligence (AI), due to developments in Large Language Models (LLMs) (source) such as the Generative Pretrained Transformer (GPT) model (source), an opportunity has emerged to improve our healthcare. The text prediction capabilities of LLMs enable it to generate high-quality text, based on natural language inputs. Bringing on potential use cases for assisting healthcare professionals, such as dieticians and therapists.

One of these use cases is the generation of motivational text messages to support adherence to, and the effectiveness of behaviour change interventions, both during and after. This paper covers design research that explores how we can develop a plenary set of validated motivational (SMS) text messages via GPT 4 according to the Behaviour Change Wheel framework to support the reinforcement of healthy dietary behaviours.

The study compares the performance of GPT-generated messages to a selection of handcrafted and validated messages from a study in the Society of Behavioural Medicine journal by MacPherson et al. (2021c). The performance of these messages is evaluated by Dutch dieticians on a decomposition of the APEASE model, a model which is commonly used to evaluate behaviour change interventions from the Behaviour Change Wheel (BCW) theory

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

TU/e, June, 2023, Eindhoven, Netherlands

© 2018 Copyright held by the owner/author(s).

(Michie et al. 2014). The prompt used to initiate the message generation by GPT was designed to generate motivational text messages for the same message length, target behaviour, and behaviour change technique using the definitions found in the BCW theory.

KEYWORDS

Large Language Models, GPT-4, Behaviour Change, Obesity, Dietary Interventions, Intervention Aftercare,

ACM Reference format:

Joris Raaphorst. 2023. Surname, FirstName Surname and FirstName Surname. 2018. Generating Motivation for Obesity Prevention: Reinforcing behaviour change with GPT-4 generated motivational text messages. *Eindhoven, Netherlands*.

Reader Guide

In the introduction obesity prevention, the emergence of AI and the workings of GPT are covered. Followed by an overview of the research process, which consists of a literature review, design process and research set-up. After the results are covered, followed by the data analyses, conclusion, recommendations, discussion, and finally some acknowledgments. To keep this paper within a 5000-word limit, parts are minimised, to learn more about any topic, please consult the Appendix, in which many summarizations are explained.

1. Introduction

1.1 Behaviour change interventions for obesity

That healthcare is both viewed as important as it is costly, is exemplified by the Dutch government's budget allocation, spending 26,8% of their total budget on healthcare, making it their largest spending post (Ministerie van VWS, 2022).

Obesity, a condition that is influenced by the complex interplay of behavioural, environmental, and genetic factors (Rosen, 2014), severely increases morbidity, and is associated with diseases like; cardiovascular disease; type 2 diabetes; musculoskeletal disorders, and cancer (World Health Organization, 2023). In more than 70 countries the obesity rates have doubled since 1980 (Collaborators GBDO, 2017). According to the Dutch CBS (Centraal Bureau Statistiek, 2021) the age group of 18-30 is with 30% of the people being overweight, the best-performing group of all the adults

(Appendix A1.1). Resulting in a large sum of Dutch citizens with health risks that tax the healthcare system.

Losing body weight is a matter of a lower energy intake than the energy consumption of the body (Hill et al. 2020). As such obesity prevention generally covers two behaviours; dietary intake and physical activity (PA) (Shinall et al., 2021). This study focuses on dieticians, helping people with analysing and optimising their nutritional intake, and playing a role in establishing healthy eating behaviours (U.S. Bureau of Labor, 2022). The interaction between healthcare professionals and behaviour change subjects is also referred to as an intervention, which is generally carefully designed by behaviour change experts.

A commonly used outline for these designs is the Behaviour Change Wheel (BCW) framework, which is a synthesis of 19 behaviour change frameworks. It follows an eight-step process and introduces several models such as COM-B (Capability, Opportunity, Motivation); the TDF (theoretical domains framework) for composing a behaviour change intervention (Appendix A2.1.) and the APEASE criteria for evaluation of the effectiveness of an intervention (Appendix A2.) (Michie et al. 2011; Michie et al. 2014).

By following the BCW process a wide variety of interventions can be designed. In the study by MacPherson et al. (2021c), the BCW is applied, resulting in an intervention strategy for a text message-based intervention and “124 theory-based diet & physical activity text messages”. These Motivational Text Messages (MTMs) were carefully handcrafted and iteratively evaluated and validated according to the APEASE criteria. These messages are being used by professionals as during and after interventions to support patients in the desired behaviour change actions.

Research shows that providing consistent aftercare in preventive interventions is desirable as it increases the reinforcement of the achieved behaviour change, improving the effectiveness of behaviour change intervention (Kwasnicka et al. 2016). Given the large diversity of the people, it is expected that more personalised and tailored messages are desired to fit into the specific personal situations of individuals.

However, due to the growing global shortage of healthcare professionals (WHO, 2019) providing consistent and personalised aftercare and hardly feasible. This study considers this as an interesting use case for the emerging technology of Large Language Models (LLMs) such as GPT due to their capacity to quickly generate high quality texts with low costs.

1.2 The rise of Artificial Intelligence

These related works are extended in Appendix A3.1.

The exponential acceleration in AI technology and the emergence of OpenAI’s GPT technology is not sudden, as it has been a topic of research since 1943 (McCulloch & Pitts, 1943). AI research is marked by two AI winters, periods in which hardly any research

had been performed due to cutbacks on funding, during the period of 1970-1979 and 1987-1993.

That GPT technology is rapidly developing speaks from examples such as, OpenAI’s release of GPT-4, their introduction of plugins and the emergence of numerous open-source initiatives on the Machine learning repository HuggingFace platform such as AutoGPT. With Microsoft investing billions of dollars into this technology (Microsoft, 2023), strong and fast development is to be expected in the upcoming years, and the impact on people is predicted to be tremendous (Rotman, 2023).

The promise of the GPT technology has a high potential to innovate domains across the whole industry. Studies are starting to show that GPT technology is capable of outperforming doctors in some use cases such as replying to patient messages (Ayers et al., 2023). When healthcare professionals are supported by AI technology in text generation tasks, they might be able to help more people personally. Yet, a petition named the “Pause Giant AI Experiments: An Open Letter” (Future of Life Institute, March 2023) signed by many prominent tech innovators, calls to slow down our progress and to first develop safeguards and regulation (Future of Life Institute, 2023a). The goal of this study is therefore to create a deeper understanding of the potential benefits, limits and risk of using GPT-4 for reinforcing dietary behaviour change via MTM’s.

1.3 Generative Pre-trained Transformers (GPT)

These related works are extended in Appendix A3.2 - A3.4

GPT is an autoregressive, neural network, machine learning model, trained on data from the internet using Natural Language Processing (NLP). The stimulus or input used to task an LLM is called a prompt. The discipline of creating a prompt which results in the desired output is prompt engineering (Appendix 3.2).

Literature distinguishes several forms of prompt, namely few-shot, 0-shot, and 1-shot which are used to both train and interact with a model (Brown et al., 2020). The word shot refers to the number of examples that are added to the prompt. A prompt may list a few (labelled) examples, to show what is considered a good or bad output, which the LLM then attempts to complete by generating new entries for the list. Or no examples, by explaining in natural language what action should be taken. Lastly a mixed method can be applied in which both natural language tasking and supplying examples are used to prompt the LLM (Appendix 3.3.1).

GPT technology offers, only in their playground and API, a 0-1 temperature scale which can be used to influence probability tolerance, or Markov assumption (Markov, 1945), the model makes. For this study the default setting of 0,7 is used, this setting allows for enough variation in the output, but refrains from highly improbable results. Additionally, it is the default setting of GPT and is as such what most people use when they interact with the GPT model (Appendix 3.4.1).

2. Design Research Process

This research was inspired on the theoretical works from MacPherson et al. (2021). The study was conducted following three processes which are visualised in Figure 1. Firstly a literature review was conducted to identify the gap in research. Secondly, the ideation phase to design the prompt for GPT to generate MTMs. These two phases co-evolved as new literature was becoming available during the ideation phase. Thirdly, three surveys were developed to invite dieticians to evaluate these messages.

2.1 Literature review

For the literature review Google Scholar was utilised to identify relevant literature, as well as GPT-4. Due to the natural language understanding of GPT-4 and its excess to the internet and plugins such as Xpapers and ScholarAI it was able to identify literature that held relevance but did not comply with the direct search query. The three main topics that were investigated for which the queries iteratively evolved. The topics covered; Obesity and overweight and its relations to lifestyle and its relation to morbidity; BCW-theory and its application in several use cases; The emergence of AI and the workings of LLMs such as GPT and Prompt Engineering with GPT-4. A summary of the insights of the literature review is covered in the Related Works section, which are extended on in the Appendix (A1. /A3.)

2.3 Design research probe design & prototype

To elicit the ‘potential benefits, limits, and risk of GPT technology within the context of dietary behaviour change interventions’ the generated MTM’s had to be as similar as possible to the handcrafted MTM’s while drawing as much as possible on the neural network of GPT-4. Meaning that the prompt had to generate: ‘One motivational text message; no longer than 160 characters; Which motivates the receiver to improve their dietary target behaviour; and is semantically as comparable as possible to the selected handcrafted MTM’s (Appendix A6.1) from the MacPherson case study, by using the same behaviour change technique (Appendix A1.2).’ An iterative, six step, design process resulted in a prompt which only supplies the model with the same definitions from the BCW theory, as used for the handcrafted messages.

3. Data collection

3.1 Candidate messages

The handcrafted MTMs from the case study from MacPherson et al., (2021c) were designed based on 28 BCTs, leaving us with roughly 4,5 messages per BCT. To take some of the variety between these BCT’s into account, one message from three different BCT’s (Social Support, Goal Setting, and Self-Reward), covering the same TB, were selected for benchmarking (Appendix A6.1). No specific method was used for the selection, except that all three are commonly used BCT’s which vary widely in nature.

By supplying the core prompt with one of the handcrafted MTM’s as example (one-shot) and its corresponding BCT description, three

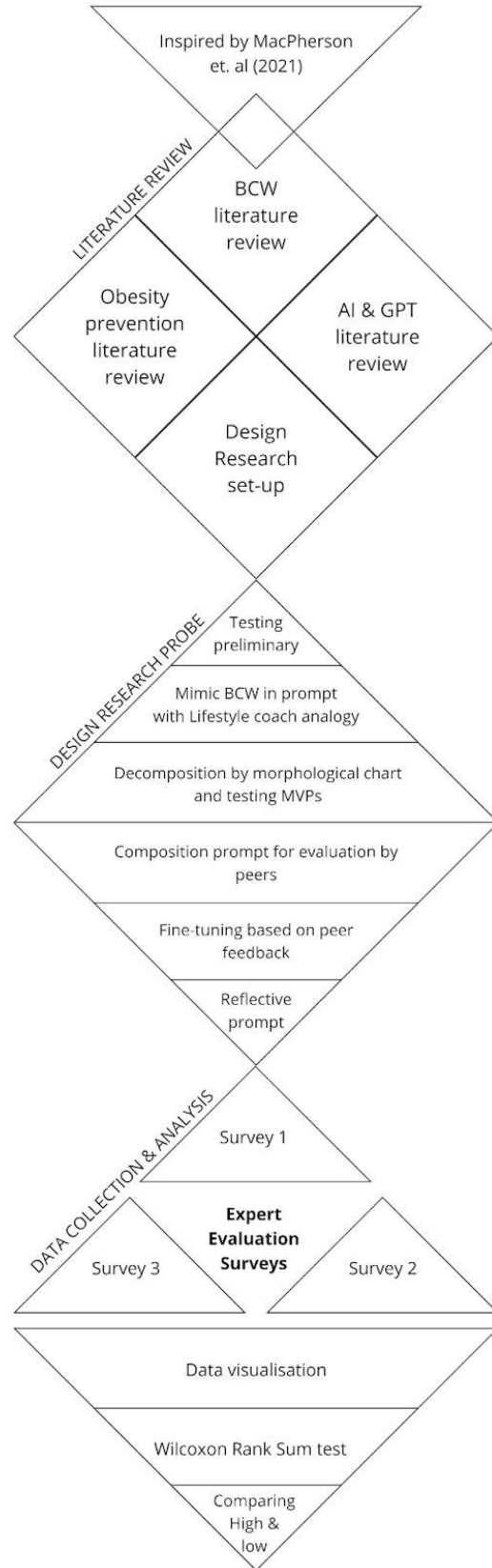


Table 1 - Design Research overview

different prompts were created (Appendix A6.2). Making the handcrafted MTM's parent to the generated MTM. GPT-4 was prompted five times for every parent message, each time in a new conversation (required since previously generated messages affect its new outputs. This resulted in a total of 18 messages, 3 parents and 15 children (Appendix A6.3).

3.2 Participants

To mitigate 'survey fatigue' which compromises the reliability of the results and causes nonrespondents (Porter et. al., 2004). The MTM's are distributed in three separate surveys, each containing one parent MTM and its five corresponding child MTMs. All three surveys were distributed among 250 dieticians registered as proficient in treating obesity at the NVD (Nederlandse Vereniging van Dietisten). No other characteristics (e.g. gender, age, experience) were taken into account for this study. The target response rate, to receive strong evidence, was 25 recipients per survey, assuming a 10 percent response per survey.

3.3 Scale distribution

For each MTM, data was collected via a twelve item, five-point Likert scale (Likert, 1932). Which is by many conceived as one of the most user-friendly of scales (Van Learhoven, 2004), being easy and reliable (Royeen, 1985) allowing for simple measurement of participant perception (Subedi, 2016). The study uses a 5-point scale, because the difference in reliability versus 7, 9 or 11-point scales is negligible (Croasmun & Ostrom, 2011), but is reducing the cognitive load for participants (Lee & Pek, 2014).

Likert scale surveys result in ordinal data, however, if the scale is linguistically symmetrical and revolves around a neutral midpoint, the distance between the points can be assumed as equal. (Norman, 2010). As such a common point designation is used; 1) Strongly disagree 2) Disagree 3) Neutral 4) Agree and 5) Strongly Agree. Or in Dutch (since this study was performed among a Dutch community of dieticians) 1) Helemaal mee oneens 2) Mee oneens 3) Neutraal 4) Mee eens 5) Helemaal mee oneens.

3.4 Evaluation criteria

The parent messages were evaluated and validated using the APEASE criteria, developed to assist intervention designers in selecting and evaluating intervention functions, policy categories, behaviour change techniques and the eventual intervention strategy (Michie et al, 2014). Each criterion is provided with a definition in which its interpretation is explained (Appendix 6.4), however no detailed specification appears to be provided about how they should be used for evaluation.

Reviewing studies that apply the model shows that APEASE is applied in a variety of ways. Examples can be found of studies using the criteria in a survey (Ginis et. al, 2020) and via expert panels (Crayton et. al, 2018; (MacPherson et. al, 2021c). In some cases, the direct definition is used (Scott et. al, 2021), in other examples the definition is modified to suit the context (Atkins, 2016) or decomposed, increasing the granularity of the collected

insights (Forbes et. al, 2022) and improving the latency of the variables (Bollen, 2020).

The APEASE descriptions consider both the effectiveness of the *delivery* and the *implementation* of the intervention. Since this study investigates the implementation effect of the message content, and not the delivery of such a message, a modified version of the APEASE definitions is used. Next to this each definition is decomposed into two Criteria Statements (Appendix A6.5). For this, an equal distribution of factors that embody semantically related, tangible phenomena (Harpe, 2015) was kept in mind. To format the Criteria Statements in a Microsoft Likert scale form (selected for data protection reasons) the statements had to be reformulated to minimise their length in the table (Table 1). The messages were then translated into Dutch (nationality of the research population) using back translation (Bennet, 2022).

3.4.1 Modified APEASE criteria statements

The text message (or its implementation) does not require;

Affordability	Financial costs (explicit or implicit)
	High linguistic proficiency
Practicability	Physical, social or financial resources
	Knowledge or skill

The text message (or its implementation) is;

Effectiveness	Cost-effective for behavioural change
	Effort effective for behavioural change
Acceptability	Appropriate for those with disabilities (neurodiverse/ physical/ mental)
	Appropriate for minorities (ethnic/ cultural/ religious/ sexual/ gender)

The text message (or its implementation) does not cause;

Side-effects/Safety	Unsafe behaviour
	Unintended side effects
Equity	Inequality in standard of living
	Inequality in health or well-being

Table 2 - Modified APEASE Criteria Statement

4. Result & analysis

4.1 The final prompt design

To achieve the best comparability to the baseline MTMs a minimal prompt is used, only providing a role, task, TB, BCT name and description (because it is not part of its neural network) and one example message (Table 2). Within the same prompt it is tasked to perform a fact check and to reflect on itself within a table format (Table 3).

Message generation prompt	
Direct task specification with dietician analogy	Act as a dietician, write one motivational text message that a dietician would create.
Format constraints	Use a maximum of 160 characters.
Target behaviour	I would like to achieve the improvement of dietary habits
BTC	According to the behaviour change technique: 'BCT NAME PLACEHOLDER'
BTC description	This behaviour change technique is defined as follows: 'BCT DESCRIPTION PLACEHOLDER'
Specification by baseline messages	Comparable to:
	"EXAMPLE MESSAGE PLACEHOLDER"
	[...]

Table 3 - Message generation prompt used in Research Probe

Reflective Prompt	
Fact Check List	Generate a set of facts that are contained in the 'Motivational text message' you have generated. The set of facts should be inserted in a specific point in the 'Motivational text message'. The set of facts should be the fundamental facts that could undermine the veracity of the 'Motivational text message' if any of them are incorrect.
Reflection	Whenever you generate an 'Motivational text message' Explain the reasoning and assumptions behind your 'Motivational text message'.
Format structure	Create a table:
	Motivational Text Message Facts Reasoning Assumptions

	"Message" 1) Fact explanation 2) Fact explanation 3) etc. Output Reasoning Output Assumptions

Table 4 - Reflective Prompt used in Research Probe

4.2 Respondents

The NVD was kind enough to place an advertisement, requesting participation in their biweekly newsletter. Additionally, 250 Dutch dieticians were addressed via email addresses acquired from the NVD. All these recipients were labelled by the NVD to be proficient in dealing with obesity. From these respondents no further data was collected (e.g. gender, experience, age, etc.)

Survey number	Behaviour change technique	Respondents
1	Social Support	18
2	Goal Setting	10
3	Self-Reward	11

Table 5 - Respondents information

The response was lower than anticipated (Table 4), various reasons might have caused this, some feedback from the respondents point towards; survey fatigue; failing to convey the relevance; or a daunting ethics and privacy conduct, among others.

4.3 Evaluation results

For the interpretation of Likert scale data, the data was converted to a numerical format (1=Totally Disagree & 5=Totally Agree). Then, for each criteria statement then a mean could be calculated (Table 5). In the table the grey - B represents the baseline message, and the others the generated messages.

MEANS	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
M1 - B	3,3	3,1	2,7	2,8	3,4	3,6	3,2	3,3	3,5	3,3	3,5	3,4
M1.1	3,4	3,1	2,6	2,3	3,2	2,8	2,8	2,9	3,1	2,7	2,9	3,0
M1.2	3,4	3,2	2,3	2,3	3,3	3,1	2,9	3,0	2,6	2,5	2,6	2,6
M1.3	3,3	2,6	2,3	2,4	3,6	3,2	2,9	3,2	3,0	2,7	3,0	2,9
M1.4	3,3	3,1	2,7	2,6	3,5	3,4	3,2	3,3	3,2	2,9	3,1	3,1
M1.5	3,4	2,8	2,7	2,5	3,6	3,3	3,0	3,3	3,2	3,1	3,1	3,2
M2 - B	3,4	3,1	3,1	2,1	3,7	3,4	2,8	3,3	3,5	3,2	3,3	3,2
M2.1	3,4	3,2	3,1	2,4	3,8	3,5	3,0	3,4	3,5	3,2	3,5	3,5
M2.2	3,2	3,3	3,0	2,6	3,3	3,3	3,0	3,3	4,1	3,8	3,7	3,6
M2.3	2,8	3,3	2,9	2,8	3,3	3,3	3,5	3,7	3,8	3,4	3,6	3,7
M2.4	3,6	3,7	3,5	3,6	3,7	3,6	3,6	3,9	3,8	3,8	4,0	4,0
M2.5	3,2	2,5	2,8	1,8	3,4	3,1	2,7	3,3	3,7	3,4	3,6	3,6
M3 - B	2,9	2,7	2,7	2,1	3,2	2,9	2,9	3,4	3,5	3,0	3,3	3,5
M3.1	2,8	2,9	2,9	2,5	3,1	3,5	3,0	3,2	3,3	2,2	3,0	3,1
M3.2	3,1	3,1	3,1	2,4	3,2	3,2	2,7	3,1	3,1	2,5	2,8	3,1
M3.3	3,5	3,8	3,7	3,5	3,5	3,5	3,7	4,0	4,0	3,9	4,0	4,0
M3.4	3,1	3,0	3,3	2,8	3,2	3,2	3,4	3,4	3,2	2,9	3,4	3,5
M3.5	2,8	2,9	2,9	2,5	3,1	3,5	3,0	3,2	3,3	2,2	3,0	3,1

Table 6 - Evaluation results, Mean per Criteria Statement

5. Analyses

5.1 Data visualisation

Each criteria statement embodies a different, independent factor from the APEASE criteria. They are interpreted individually and not combined to calculate a composite mean for a whole criterion (e.g. A1+A2) or the overall intervention effectiveness (e.g. all twelve criteria statements). When we accumulate the means of the generated messages from a survey and plot them onto a graph, a pattern emerges (Figure 1, 2 & 3).



Figure 2 - Survey 1 - Social Support (Mean minus 3)



Figure 3 - Survey 2 - BCT: Goal Setting (Mean minus 3)



Figure 4 - Survey 3 - BCT: Self-Reward (Mean minus 3)

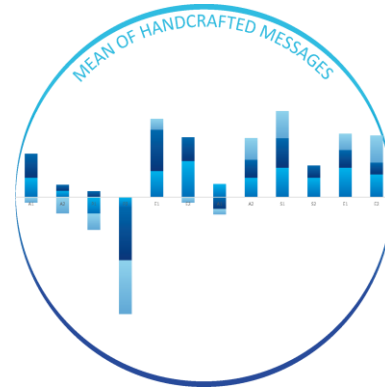


Figure 5 - Performance Baseline MTMs (Mean minus 3)

This same pattern seems to emerge when we accumulate the means of the handcrafted messages (Figure 5), which is probably due to the nature of delivering SMS-like messages for dietary behaviour change purposes. To compensate for this, the scores of the values of the handcrafted messages are used as a baseline, by deducting the values of one handcrafted message from its corresponding generated messages (Figure 6).



Figure 6 - Survey 1, 2 & 3 - Compensated by Baseline

Figure 6 shows that the BCT ‘social support’ is underperforming in comparison to the BCT ‘goal setting’ and ‘self-reward’. And even though the handcrafted message from the BCT ‘social support’ is slightly outperforming the handcrafted message from test 2 and 3. The non-compensated means (Figure 1, 2 & 3) clearly show that the generated messages of test 1 are outperformed by the generated messages from test 2 and 3.

5.2 Wilcoxon Rank Sum Test

In order to elicit the extremes among the messages, the means of all criteria per message were summed and then ranked. Surprisingly neither the highest five or the lowest five scoring messages contain any handcrafted messages (Table 6).

Rank 1 M3.3	Rank 7 M3.4	Rank 13 M3.1
Rank 2 M2.4	Rank 8 Baseline M2	Rank 14 M3.5
Rank 3 M2.2	Rank 9 M1.4	Rank 15 M3.2
Rank 4 M2.3	Rank 10 M1.5	Rank 16 M1.3
Rank 5 M2.1	Rank 11 M2.5	Rank 17 M1.1
Rank 6 Baseline M1	Rank 12 Baseline M3	Rank 18 M1.2

Table 7 - MTM ranking based on sum of means.

Due to a lower anticipated response rate, and the ordinal nature of the data, careful consideration is required for making assumptions about the data. Based on the visualised data, using histograms, (APPENDIX X) we concluded that the assumptions for interval or normal distribution and homogeneity of variance are not being met. As such a two-tailed non-parametric Wilcoxon-Rank-Sum test is used to test the strength of the evidence according to the following Null Hypothesis (H0): ‘The mean of a Criteria Statement for the AI-generated message is equal to the mean of a Criteria Statement for the baseline message’.

$$H_0 = \mu_{AI} = \mu_{Baseline} \quad \& \quad H_a = \mu_{AI} \neq \mu_{Baseline}$$

	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
M1.1	0,89	0,87	0,72	0,15	0,43	0,04	0,28	0,31	0,26	0,11	0,05	0,25
M1.2	0,85	0,67	0,25	0,25	0,73	0,11	0,48	0,30	0,03	0,03	0,01	0,03
M1.3	0,89	0,12	0,19	0,23	0,76	0,25	0,41	0,72	0,17	0,14	0,16	0,19
M1.4	0,89	0,87	0,97	0,56	0,99	0,58	0,84	1,00	0,33	0,25	0,09	0,30
M1.5	0,89	0,39	1,00	0,41	0,67	0,39	0,56	1,00	0,55	0,57	0,15	0,46
M2.1	1,00	0,82	1,00	0,85	0,85	0,85	0,76	0,82	0,97	1,00	0,76	0,60
M2.2	0,76	0,60	1,00	0,31	0,52	0,82	0,73	0,97	0,20	0,36	0,36	0,36
M2.3	0,33	0,71	0,73	0,27	0,43	0,85	0,34	0,36	0,55	0,79	0,57	0,85
M2.4	0,76	0,31	0,55	0,02	1,00	0,76	0,27	0,29	0,45	0,34	0,14	0,11
M2.5	0,71	0,27	0,57	0,47	0,57	0,60	0,91	1,00	0,68	0,76	0,57	0,45
M3.1	0,92	0,74	1,00	0,45	0,60	0,28	0,74	0,49	0,43	0,24	0,74	0,55
M3.2	0,60	0,60	0,60	0,20	0,49	0,60	0,77	0,55	0,67	0,49	0,90	0,55
M3.3	0,31	0,10	0,12	0,31	0,24	0,24	0,12	0,05	0,07	0,08	0,05	0,05
M3.4	0,72	0,79	0,38	0,84	0,53	0,60	0,32	0,32	0,72	0,92	0,32	0,28
M3.5	0,60	0,29	0,74	0,84	0,34	0,39	0,32	0,49	0,29	0,69	0,21	0,38

Table 8 - P-values per Criteria Statement in relation to the baseline mean

The resulting p-values (Table 7) are colour coded for the following alpha levels: 0,05;0,1;0,2;0,3. And show that both for the two top ranking (M3.3 & M2.4) and the two lowest ranking messages (M1.1 & M1.2) strong enough evidence was found to reject the NULL-hypothesis using the 0,05 alpha level. For each message a detailed and visual overview can be found in Appendix 8.1.

5.3 Comparison of extremes

5.3.1 Underperformers

Why is GPT-4 having a harder time to write messages for the BCT Social-Support? The largest differences are found in the criteria statements which state that the message will not cause unsafe behaviour, unwanted side-effects, inequality in standard of living or health and wellbeing in comparison to the baseline message: “Sometimes we need the support of others to help us stick to our goals. Think about who in your life can help you stick to your exercise plan.” - Baseline message from Social Support survey

By looking at the comments of the dieticians and the reflection of GPT on itself we can get an idea as to why these differences occurred (Table 8). The comments from M1.3 were also incorporated, which is justified since it seems to be part of the same trend.

M1.1	M1.2
" Lean on loved ones for healthy meal ideas & support! Together, we can build better habits."	"Rely on friends and family for healthy meal inspiration and motivation. Together, we can make positive changes."
Reflection of GPT on its assumptions	
“The recipient's social circle has knowledge of healthy meal ideas”	
Dietician comments:	
<p>M1.1 (6 dieticians commented something comparable) “unhealthy foods are associated with unhealthy environment, so leaning on loved ones for help is good but now the suggestion is that they will know what healthy is and that is often not true”</p>	
<p>M1.3 (3 dieticians commented something comparable) “The problem is often the lack of a supportive network with unhealthy eating habits, that's the problem that should be solved. not sharing mealplans and growing together”</p>	

Table 9 - Relevant comments of participants and reflection by GPT-4

According to the respondents, when people are obese or overweight, they do not always have access to a healthy social environment or have access to social support at all. However, GPT-4 consistently assumed that receivers have access to a healthy social environment. This resulted in a low performance on both the *Safety/Side-effect* and *Equity* criteria. These low scoring criteria might also have impacted the low performance on *Effort-Effectiveness*. No further evidence was found to explain this difference.

5.3.2 High-performers

GPT-4's performance on the social support messages is in stark contrast with its performance on M2.4 and M3.3. Both these messages score consistently higher than 3 on the mean and outperform the baseline on every criterion. These messages are evidence that, according to the scale from this study, GPT-4 can generate high quality messages, outperforming specialised academics. So, why did these messages perform so well?

M2.4

"Set a goal to drink more water daily. Small habits make a big difference in the long run!".

This message performed well because "It's **about behaviour**, it's not one-way traffic information and advice". Next to this the message "suggests a specific, achievable goal" that is simple, relatively harmless, and highly accessible. Explaining the significant differences in P2 - not requiring knowledge or skill, A1 & A2 - appropriate for disabled and minorities, and E1 & E2 - not causing inequality of living and health or wellbeing.

M3.3

"Each day, reward another step towards your goals. Today, reward yourself for eating healthy. Tomorrow, reward yourself for cooking that healthy meal!".

Has a high performance because the message is non-specific, leaving space for the dietician to "eerst helpen healthy choices helder formuleren"* making it "**op maat per persoon****". Additionally, it is suggesting a mental reward, which has no chance to result in unhealthy behaviour, requires no resources, knowledge, or skills of any kind, making it an accessible option for anyone. Due to these reasons, this message has the highest score of all tested messages.

*" first help formulate healthy choices clearly".

**" custom per person".

6. Conclusion

The results prove that GPT-4 can outperform the baseline, making it a powerful tool to be used by dieticians. GPT-4's assumption that recipients have access to a healthy social support network is evidence that it is struggling to empathise with the situation of the recipients, revealing the value and necessity of human supervision. This example shows that some BCTs require additional guidance because empathy is required to align with the situation of the recipient, it appears that some BCTs are more sensitive to this than others.

7. Limitations

The limits of this study are, its prompt design and the timing, the selection of criteria, BCTs and baseline messages, the number of generated messages, the number and the identity respondents, and the strategy for analyses. The prompt, designed to be neutral, relies as much as possible on the Neural Network of the LLM. It is limited since it is still contextually framed due to word choice and its

structure. Because the LLMs neural network is dynamic and continuously learning, the timing of generating these messages influenced its output.

The choice to use a modified version of APEASE, and the selected BCTs and the corresponding baseline messages are restricting the span of the study. Other criteria, BCTs or messages might yield different results and will likely also elicit valuable insights.

The limited use of 5 generated messages per BCT was required to keep the surveys manageable and prevent fatigue, however more messages would result in better saturation, which would strengthen the evidence. For the same reason a larger sample size would also be desirable. Additionally testing with other experts (e.g. life coaches, behaviour change experts), or recipients would be valuable, but was not achievable within the constraints of this study.

The analysis assumes that lower scoring MTMs are less effective. Using the practicability criteria statements as example; 'does not require physical, social or financial resources or knowledge and skill', it can be debatable that the messages are less effective when some resources or knowledge is required.

9. Discussion

On the question, is GPT-4 capable of generating high performance MTM's? The answer is 'Yes, but...' the technology of conversational AI is very powerful and can be highly convincing, however, it is not always aligned with its users and might use wrong facts or make false assumptions. These can cause serious harm and human supervision is therefore always required.

Using the recommendations from this study, it is highly likely that the performance of GPT-4 will improve. For future works, testing the other 90 BCTs might prove very helpful in mitigating the pitfalls and sensitivity areas surrounding them. Next to this the consistency of the model could use further investigation, how does the model perform when the messages are generated over a time span? Another research direction is using the recommendations from this study to create an 'optimal prompt' to see how well it performs against the handcrafted messages from the MacPherson study. Valuable insight could be found also when co-creating messages with recipients or dieticians, eliciting their needs on a qualitative level, which could lead insights which can in turn be used to improve the prompt.

When we speculate about a point in time where a refined prompt, or new AI model, is capable of consistently outperforming the handcrafted messages in large scale studies. An interesting for some, ominous for others, future arises in which autonomous AI agents are highly equipped to effectively support personalised healthy behaviour change. How and where would we apply this technology? And will we ever reach this point in time? Or is it closer than we think, and will the first open-source AI therapists have a huge impact on the field of healthcare?

10. Recommendations

The neutral prompt used for this study utilises 93 BCTs, making it versatile in its application, and a good starting point for improvements. GPT-4 can generate nearly any sort of text; therefore, it helps to specify both what is and what is not desired outcome. Before using any of the recommendations, make sure to never put personal information into the model, GPT trains itself on the interaction with its users, and is therefore sensitive to privacy incidents!

In this chapter we provide some suggestions which are likely to improve GPT-4s performance for writing MTMs:

Including the APEASE criteria in prompt can improve the adherence to them.

1. Tasking GPT to reflect on its generated messages using the APEASE criteria might elicit sensitivity issues.
2. The BCW theory provides 7 questions which help to frame the context of an intervention, using these has shown to result in interesting and more appropriate results.
3. Explain the interaction between recipient and dietician, allowing GPT-4 to take the role of the dietician into account (like in M3.3).
4. Providing specific goals and tasks for the recipient will prevent GPT-4 from writing prompts which require recipients to formulate their own path towards behaviour change, (like in M2.4).
5. Providing the model with good examples allows the model to use pattern recognition, which has proven to improve the results, both quantity and quality matter.
6. Continuously iterate based on the responses from your clients, each time improving on the prompt, working towards perfection.

Acknowledgements

For this study first and foremost I want to acknowledge the support of Associate Professor Lu Yuan, who challenged both the study and me as researcher. Additionally, I want to thank the NVD, as they shared the surveys among their members via their newsletter. Lastly, the role of GPT-4, it was used for generating the MTM's, and to support literature exploration. It suggested papers, helped to formulate search queries and helped me to understand some complex technical concepts. The model was not used to generate any text for this paper.

References

Contains references from both this paper and its appendix.

Atkins, L. (2016). Using the Behaviour Change Wheel in infection prevention and control practice. *Journal of Infection Prevention*, 17(2), 74–78.

<https://doi.org/10.1177/1757177415615952>

CBS. (2022, 9 september). Hoeveel volwassenen hebben overgewicht? - Nederland in cijfers 2022. Hoeveel volwassenen hebben overgewicht? - Nederland in cijfers 2022 | CBS. <https://longreads.cbs.nl/nederland-in-cijfers-2022/hoeveel-volwassenen-hebben-overgewicht/>

Chan, A. (2022). GPT-3 and InstructGPT: technological dystopianism, utopianism, and “Contextual” perspectives in AI ethics and industry. *AI and Ethics*, 1-12.

Church, K., & Liberman, M. (2021). The future of computational linguistics: On beyond alchemy. *Frontiers in Artificial Intelligence*, 4, 625341.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21-27.

Crayton, E., Wright, A., & Ashworth, M. (2018). Improving medication adherence in stroke survivors: the intervention development process. *BMC Health Services Research*, 18(1). <https://doi.org/10.1186/s12913-018-3572-1>

Czinczoll, T., Yannakoudakis, H., Mishra, P., & Shutova, E. (2022). Scientific and Creative Analogies in Pretrained Language Models. *arXiv preprint arXiv:2211.15268*.

Dan Hendrycks, Nicholas Carlini, John Schulman, and Jacob Steinhardt. Unsolved problems in ml safety. *ArXiv*, 2021b. URL <https://arxiv.org/abs/2109.13916>.

Davidson, K. W., & Scholz, U. (2020). Understanding and predicting health behaviour change: a contemporary view through the lenses of meta-reviews. *Health Psychology Review*, 14(1), 1–5. <https://doi.org/10.1080/17437199.2020.1719368>

Dean, J., Corrado, G., Monga, R., Chen, K., Devin, M., Mao, M., ... & Ng, A. (2012). Large scale distributed deep networks. *Advances in neural information processing systems*, 25.

DeJong, G., & Mooney, R. (1986). Explanation-based learning: An alternative view. *Machine learning*, 1, 145-176.

E. Gamma, R. Johnson, R. Helm, R. E. Johnson, and J. Vlissides, *Design patterns: elements of reusable object-oriented software*. Pearson Deutschland GmbH, 1995.

Fabio Petroni, Tim Rocktaschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H. Miller, and Sebastian Riedel. Language models as knowledge bases? In *EMNLP*, 2019.

Forbes, C., Morley, N., Liabo, K., Bjornstad, G., Boulton, H., Ahmed, S. U., Ciesla, K., Vafai, Y., Bridges, S., Logan, S., & Berry, V. L. (2022). Prioritising child health and maternity evidence-based interventions or service models: a stakeholder-driven process. *BMC Health Services Research*, 22(1). <https://doi.org/10.1186/s12913-022-08110-2>

Frank, R. (1957). The perceptron, a perceiving and recognizing automaton Project Para. Cornell Aeronautical Laboratory.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4), 193-202.

Ginis, K. A. M., Cheifetz, O., Todd, K. R., Chebaro, C., Phang, S. H., Shaw, R. A., & Whaley, K. J. (2020). Co-development of a physiotherapist-delivered physical activity intervention for adults with spinal cord injury. *Spinal Cord*, 58(7), 778–786. <https://doi.org/10.1038/s41393-020-0422-x>

Goodfellow, I. J. (2014). On distinguishability criteria for estimating generative models. *arXiv preprint arXiv:1412.6515*.

Gwern Branwen. “GPT-3 Creative Fiction”. In: (2020).

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, 313(5786), 504-507.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685-695.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou. Chain of thought

prompting elicits reasoning in large language models. In *NeurIPS*, 2022b.

Kann, B. H., Hosny, A., & Aerts, H. J. (2021). Artificial intelligence for clinical oncology. *Cancer Cell*, 39(7), 916-927.

Kelley, H. J. (1960). Gradient theory of optimal flight paths. *Ars Journal*, 30(10), 947-954.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, 2.

Leswing, K. (2023b, februari 15). Microsoft's Bing A.I. made several factual errors in last week's launch demo. *CNBC*. <https://www.cnbc.com/2023/02/14/microsoft-bing-ai-made-several-errors-in-launch-demo-last-week-.html>

MacPherson, M. K., Cranston, K., Locke, S. R., Bourne, J. E., & Jung, M. E. (2021c). Using the behavior change wheel to develop text messages to promote diet and physical activity adherence following a diabetes prevention program. *Translational behavioral medicine*, 11(8), 1585–1595. <https://doi.org/10.1093/tbm/ibab058>

Markov, A. A. (1954). The theory of algorithms. *Trudy Matematicheskogo Instituta Imeni VA Steklova*, 42, 3-375.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5, 115-133.

Ministerie van Volksgezondheid, Welzijn en Sport. (2022, 27 december). Zorgkosten stijgen in 2022: basispakket 2,9% en langdurige zorg 4,8%. *Nieuwsbericht | Zorginstituut Nederland*. <https://www.zorginstituutnederland.nl/actueel/nieuws/2022/12/28/zorgkosten-2022#:~:text=In%202022%20stijgen%20de%20kosten,nieuwe%20kwartaalcijfers%20van%20Zorginstituut%20Nederland.>

Norman, G. R. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15(5), 625–632. <https://doi.org/10.1007/s10459-010-9222-y>

OpenAI API. (z.d.). <https://platform.openai.com/docs/guides/completion/prompt-design>

P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, “Pretrain, prompt, and predict: A systematic survey of prompting methods in natural language processing,” *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–35, 2023.

Reynolds, L., & McDonell, K. (2021, May). Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-7).

Schapiro, R. E. (1990). The strength of weak learnability. *Machine learning*, 5, 197-227.

Schjøll Abildgaard, A. (2023, 17 februari). Large language model chatbots are prone to error. *Financial Times*. <https://www.ft.com/content/d7ab4528-ddda-420a-abf9-8e0f05da8be8>

Schon, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.

Scott, S., May, H., Patel, M., Wright, D. W., & Bhattacharya, D. (2021). A practitioner behaviour change intervention for deprescribing in the hospital setting. *Age and Ageing*, 50(2), 581–586. <https://doi.org/10.1093/ageing/afaa169>

Shaw, G. R. (1986). Donald Hebb: The Organization of Behavior. In *Springer eBooks* (pp. 231–233). Springer Nature. https://doi.org/10.1007/978-3-642-70911-1_15

Sherman, B. N. (2023, 8 februari). Google's Bard AI bot mistake wipes \$100bn off shares. *BBC News*. <https://www.bbc.com/news/business-64576225>

Shinn, Noah, Beck Labash, and Ashwin Gopinath. "Reflexion: an autonomous agent with dynamic memory and self-reflection." *arXiv preprint arXiv:2303.11366* (2023).

Si, C., Gan, Z., Yang, Z., Wang, S., Wang, J., Boyd-Graber, J., & Wang, L. (2022). Prompting gpt-3 to be reliable. *arXiv preprint arXiv:2210.09150*.

Tom B Brown et al. “Language models are few-shot learners”. In: *arXiv preprint arXiv:2005.14165* (2020).

Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX(236), 433–460. <https://doi.org/10.1093/mind/lix.236.433>

Turing, Alan M., and R. Brooker. "Programmers' handbook for the manchester electronic computer mark ii." University of Manchester (1952).

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., ... & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. arXiv preprint arXiv:2302.11382.

Yang Trista Cao, Yada Pruksachatkun, Kai-Wei Chang, Rahul Gupta, Varun Kumar, J. Dhamala, and Aram Galstyan. On the intrinsic and extrinsic fairness evaluation metrics for contextualized language representations. In ACL, 2022.

Appendix Generated Motivation

Contents

- A0. ERB & Survey questions 4
 - A0.1.1 Introduction to the survey 4
 - A0.1.2 – Information sheet (provided with the survey) 4
 - A0.1.3 – Consent questions 6
 - A0.2.1 – Survey Questions 7
- A1. Introduction 11
 - A.1.1 CBS numbers of overweight & obesity in the Netherlands 11
- A2. Behavior change wheel 12
 - A.2.1 Behaviour Change Wheel (BCW) 12
 - A.2.2. Original APEASE criteria 12
 - A.2.3 7 Situational Context questions from the BCW framework 13
- A3. AI & GPT 14
 - A.3.1 History of AI development 14
 - A3.2 Summary of GPT 15
 - A3.3 Prompt engineering 16
 - A3.3.1 Task specification 16
 - A3.3.3 Prompt Patterns 17
 - A3.4 Reliability of LLMs 18
 - A3.4.1 Generalizability 18
 - A3.4.2 Confidence calibration 18
 - A3.4.3 Social bias and fairness 18
 - A3.4.4 Factuality 19
- A4. Experimental Generative Prompts 20
 - A4.0 Process overview 20
 - A4.1 Iteration 1 21
 - A4.1.1 Test 1 21
 - A4.1.2 Test 2 21
 - A4.2 Iteration 2 22
 - A4.2.1 Test 3 22
 - A4.2.2 Test 4 23
 - A4.2.3 Test 5 24
 - A4.3 Iteration 3 26
 - A4.3.1 Morphological Chart 26

A4.3.2 Testing the behavior of separate modules.....	27
A4.4 Iteration 4.....	28
A4.4.1 Fundamental Prompt.....	28
A4.4.2 Prompt Task Specifications.....	28
A4.4.3 Test set-up.....	29
A4.5 Iteration 5.....	29
4.5.1 Fine Tuned Generative Prompt Design.....	29
A5. Evaluative Prompt design - iteration 6.....	31
A5.1 Experiment examples quantitative prompt (Review).....	31
A5.1.1 Experiment 1:.....	31
A5.1.1. 2 Experiment 2:.....	32
A5.1.1.3 Experiment 3:.....	33
A5.2 Qualitative reflective prompt.....	34
A5.2.2 Prompt design qualitative reflective prompt.....	35
A5.2.3 Output example from generative reflective prompt.....	35
A6. Research method.....	36
A6.1. Selected messages & corresponding BCT descriptions.....	36
A6.2. Final three prompts.....	36
A.6.2.1 Prompt 1 - Social Support.....	36
A.6.2.2 Prompt 2 - Goal setting.....	37
A.6.2.3 Prompt 3 - Self-Reward.....	37
A6.3. Prompt Outputs.....	38
A.6.3.1 Output prompt 1 - Social support.....	38
A.6.3.2 Output Prompt 2 - Goal setting.....	39
A6.3.3. Output Prompt 3 - Self Reward.....	40
A6.4 Criteria descriptions.....	41
A6.5 Criteria reformulations.....	42
A6.6 Criteria minimisation.....	43
A6.7 Criteria translations.....	44
A6.8 Back translations.....	45
A7. Research Method.....	46
A.8 Results.....	46
A.8.1 Baseline messages.....	46
A8.2 Generated messages.....	47
A8.3 Histograms for Assesing data distribution.....	67

A0. ERB & Survey questions

A0.1.1 Introduction to the survey

Verantwoord AI gebruik binnen de Diëtetiek (1/3)

Dit onderzoek is een initiatief van Joris Raaphorst. Het analyseert de mogelijke voordelen, risico's en beperkingen van AI (GPT-4) via een benchmark vergelijking tussen AI en specialist gegenereerde motiverende sms-berichten, bedoeld voor preventie en interventie-nazorg in de diëtik. De resultaten van dit onderzoek zijn bedoeld om een discussie op gang te brengen over de inzet van AI binnen de preventie- en interventiezorg.

De studie omvat het invullen van 3 korte enquêtes, dit is nummer 1. U beoordeelt per enquête 6 sms'jes door aan te klikken in welk mate u het eens bent met 12 korte statements, welke gebaseerd zijn op het APEASE-model.

We willen u zo min mogelijk tot last zijn en bieden daarom het onderzoek in 3 delen aan, zodat u slechts 5-8 min. per enquête nodig heeft. We hopen natuurlijk dat u alle enquêtes wilt invullen, het onderzoek is echter zo opgezet dat ook als u 1 of 2 enquêtes invult, uw resultaten kunnen worden meegenomen.

Wees ervan verzekerd dat uw antwoorden vertrouwelijk zullen worden behandeld en dat uw persoonlijke gegevens niet met derden zullen worden gedeeld. Deze studie is een initiatief van het Industrieel Design departement van de Technische Universiteit Eindhoven, en legt verantwoording af aan het Privacy team (privacy@tue.nl) en de Ethical Review Board (ethics@tue.nl) van de TU/e. Bij voorbaat dank voor het overwegen van uw deelname. Als u vragen of opmerkingen heeft, aarzel dan niet om contact met mij op te nemen via j.raaphorst@student.tue.nl.

Hieronder vindt u een informatieblad rondom privacy en ethiek, dit is gebaseerd op het standaard voorbeeld zoals aangeleverd door de TU/e. Neemt u deze alstublieft goed door! Hierna is het verlenen van toestemming een vereiste om deel te nemen aan het onderzoek. Dit proces is voor elk van de drie enquêtes exact hetzelfde.

A0.1.2 – Information sheet (provided with the survey)

U bent gevraagd om deel te nemen aan het onderzoek naar de interventie effectiviteit van GPT-4 gegenereerde berichten bedoeld voor het stimuleren van gedragsverandering binnen de diëtik ter preventie van overgewicht, omdat u geregistreerd staat als diëtist bij het NVD (Nederlandse Vereniging Diëtisten).

Deelname aan dit onderzoek is vrijwillig: u besluit zelf of u mee wilt doen. Voordat u besluit tot deelname, willen wij u vragen de volgende informatie door te lezen, zodat u weet waar het onderzoek over gaat, wat er van u verwacht wordt en hoe wij omgaan met de verwerking van uw persoonsgegevens. Op basis van die informatie kunt u middels de toestemmingsverklaring aangeven of u toestemt met deelname aan het onderzoek en met de verwerking van uw persoonsgegevens. U bent natuurlijk altijd vrij om vragen te stellen aan de onderzoeker via j.raaphorst@student.tue.nl of deze informatie te bespreken met voor u bekenden.

Doel van het onderzoek

Dit onderzoek wordt geleid door Associate Professor Lu Yuan, onderwijs directie van het Industrial Design Departement van de Technische Universiteit Eindhoven (TU/e). Het doel van dit onderzoek is om de potentiële risico's en toepassingen van de opkomende AI technologie bloot te leggen voor positieve gedragsverandering van cliënten binnen de diëtiëk.

Verwerkingsverantwoordelijke in de zin van de AVG

TU/e is verantwoordelijk voor de verwerking van uw persoonsgegevens in het kader van het onderzoek.

De contactgegevens van TU/e zijn:

Technische Universiteit Eindhoven
De Groene Loper 3
5612 AE Eindhoven

Wat houdt deelname aan de studie in?

U neemt deel aan een onderzoek waarbij we informatie zullen verzamelen via een online enquête. In deze enquête wordt u gevraagd aan te geven in welke mate u het eens bent met een twaalftal stellingen voor zes sms berichten. Voor deze studie worden drie verschillende enquêtes verstuurd, elk met een week ertussen. U kunt zelf bepalen aan hoeveel enquêtes u wilt deelnemen, al hopen we van harte dat we op uw inzet mogen rekenen!

U ontvangt voor deelname aan dit onderzoek geen vergoeding. Wel kunt u aan het einde van de enquête aangeven of u interesse hebt in de preprint van de paper die geschreven wordt naar aanleiding van dit onderzoek.

Potentiële risico's en ongemakken

Er zijn geen fysieke, juridische of economische risico's verbonden aan uw deelname aan deze studie. U hoeft geen vragen te beantwoorden die u niet wilt beantwoorden. Uw deelname is vrijwillig. Dit betekent dat u uw deelname op elk gewenst moment mag stoppen door dit te melden bij de onderzoeker. U hoeft niet uit te leggen waarom u wilt stoppen met deelname aan het onderzoek.

TOESTEMMING GEVEN

Als u op de onderstaande vragen niet comfortabel 'Ja' kunt beantwoorden kunt u helaas niet deelnemen aan ons onderzoek!

A0.1.3 – Consent questions

The following question were used to ask for permission from the participants, the participants were only able to respond with yes. The participants have to agree to all statements in order to start the survey.

1-Ik ben voldoende geïnformeerd over het onderzoek door middel van een informatieblad. Ik heb het informatieblad gelezen.

(ja)

2-Ik begrijp dat er geen andere persoonsgegevens worden gebruikt of verzameld dan de momenteel al publiekelijk openbare persoonsnaam, email-adres en professionele functie.

(ja)

3- Ik neem vrijwillig deel aan dit onderzoek. Er is geen expliciete of impliciete dwang voor mij om aan dit onderzoek deel te nemen.

(ja)

4- Ik begrijp dat alle data wordt anonimiseert voor welke andere vorm van data collectie dan ook plaats vind.

(ja)

5- De berichten zijn gegeneerd in het Engels zodat ze kunnen worden vergeleken met een benchmark studie, en een bredere relevantie dragen voor de internationale gemeenschap. Ik begrijp dat de zelfde berichten ook in het Nederlands kunnen worden gegeneerd en acht mezelf voldoende vaardig om Engels om dit te kunnen lezen.

(ja)

A0.2.1 – Survey Questions

The following question structure was used in the survey, the questions listed below cover one of the tested messages. In a survey 6 messages were covered, in total 3 surveys were conducted. This structure was designed to minimise survey fatigue. After each segment of 4 Likert scale statements the participants were given the option to leave comments.

Geef alstublieft voor elk van de 12 statements aan in welke mate u het eens of oneens bent. U doet dit voor 6 korte tekst berichten, de statements zullen voor elk bericht hetzelfde zijn. Let op, per bericht worden de 12 statements in 3 delen aangeboden, het bericht blijft dus hetzelfde!

Bericht (1/6)

"Lean on loved ones for healthy meal ideas & support! Together, we can build better habits"

Statement (1-4)

Dit bericht (of de uitvoering ervan) **vereist geen**:

	niets mee oneens	Mee oneens	Neutraal	Mee eens	Helemaal mee eens
Financiële kosten (expliciet of impliciet)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hoge taal vaardigheid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fysieke of sociale middelen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kennis of vaardigheden	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Bericht (1/6)

"Lean on loved ones for healthy meal ideas & support! Together, we can build better habits."

Statement (5-8)

Dit bericht (of de uitvoering ervan) **is**;

	Helemaal mee oneens	Mee oneens	Neutraal	Mee eens	Helemaal mee eens
Kosten- effectief t.o.v. de gedrags- verandering	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Inspannings- effectief t.o.v. de gedrags- verandering	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gepast voor beperkten (<u>neurodivers/</u> <u>ichamelijk/m</u> <u>entaal</u>)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gepast voor minderheden (<u>etnisch/cultu</u> <u>reel/religieus/</u> <u>seksueel/gen</u> <u>der</u>)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Bericht (1/6)

"Lean on loved ones for healthy meal ideas & support! Together, we can build better habits."

Statement (9-12)

Dit bericht (of de uitvoering ervan) veroorzaakt geen:

	Helemaal mee oneens	Mee oneens	Neutraal	Mee eens	Helemaal mee eens
Onveilig gedrag	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Onbedoelde bijeffecten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ongelijkheid in levens- standaarden	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ongelijkheid in Welzijn en Gezondheid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

The survey ended with the following message and left room for the participants to leave comments:

Bedankt voor uw deelname in deze studie!

Als u opmerkingen heeft of de resultaten van dit onderzoek wilt ontvangen kunt u dit hier aangeven. U kunt dit ook overslaan. U kunt dan, na het verzenden van de resultaten via **de knop onderin**, uw tabblad in de browser afsluiten.

U kunt hier een bericht achter laten voor de onderzoekers. Wat vond u van dit onderzoek? Hoe ziet u de rol van AI technologie binnen het vakgebied van de Diëtiëk? Heeft u andere op of aanmerkingen met betrekking tot het onderzoek of AI technologie?

[type here]

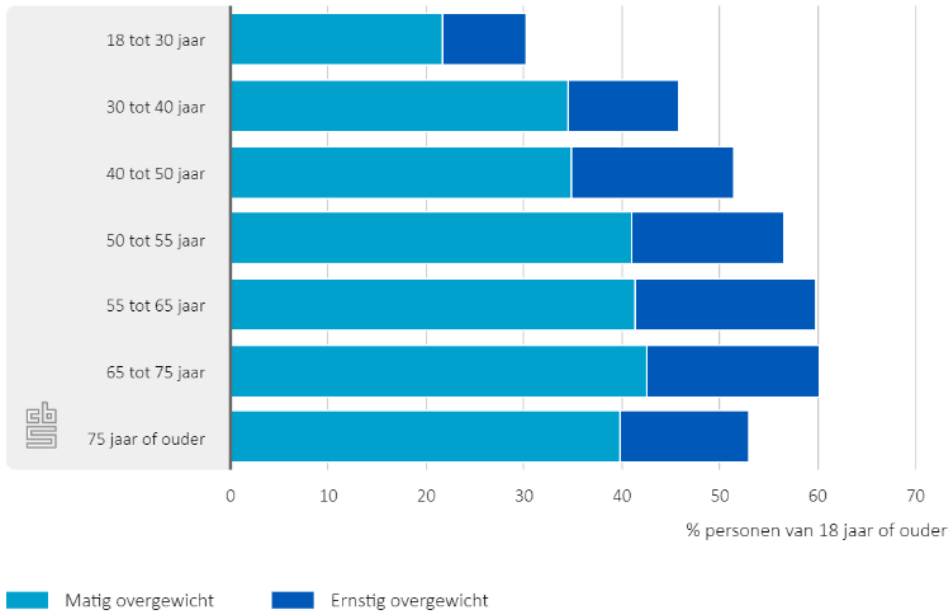
Ik ben geïnteresseerd in de ontwikkelingen van AI in het vakgebied van gedragsverandering en diëtiëk en wil graag de paper met onderzoeksresultaten ontvangen.

(ja) Ik meld me aan voor het ontvangen van de onderzoeksresultaten.

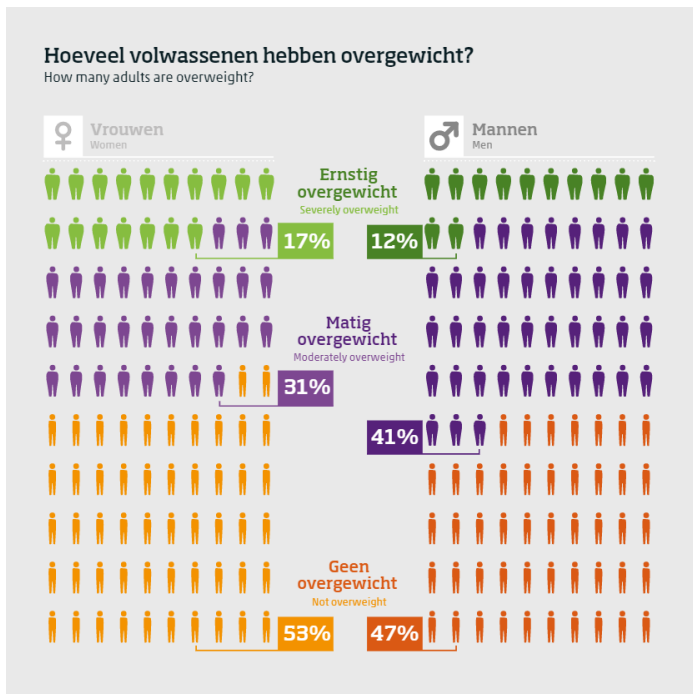
A1. Introduction

A.1.1 CBS numbers of overweight & obesity in the Netherlands

Overgewicht, 2021



Bron: CBS, RIVM



A2. Behavior change wheel

A.2.1 Behaviour Change Wheel (BCW)

The Behaviour Change Wheel model, published in 2011, and later improved on in their book publication in 2014 is a synthesis of 19 behaviour change frameworks. Providing a systematic approach for creating behaviour change interventions (fig.1).

To design a behaviour change intervention the first step is to; 1) identify the target behaviour, by creating a list of target behaviour in which the what, where and who is defined. 2) The target behaviour is selected based on impact and potential. 3) The behaviour is then further specified with questions like; What do they need to do differently to achieve the desired behaviour? Who need to perform it? With whom do they need to do it? Where will they do it? When do they need to do it?

4) The COM-B (Capability, Opportunity, Motivation) criteria are then used as a starting point to generate ideas about, what needs to happen for the target behaviour to occur? The TDF (theoretical domains framework) can be used to increase the granularity of identified changes. 5) Then a matrix is used to identify which intervention functions are applicable, which are then assessed via the APEASE model (Affordability, Practicability, Effectiveness, Acceptability, Safety and Equity). 6) The selected intervention functions can then be used to identify the policy categories via a selection matrix, and the APEASE model. 7) The selected intervention functions of theoretical domains will then guide you

through the selection of the 93 Behaviour change techniques (BCT), which are the smallest entities from which an intervention is designed. Also here an APEASE model evaluation should be applied. 8) Finally a mode of delivery, or Intervention strategy is being defined based on the outputs of all the previous steps.

A.2.2. Original APEASE criteria

Affordability

Interventions often have an implicit or explicit budget. It does not matter how effective, or even cost-effective it may be if it cannot be afforded. An intervention is affordable if within an acceptable budget it can be delivered to, or accessed by, all those for whom it would be relevant or of benefit.

Practicability

An intervention is practicable to the extent that it can be delivered as designed through the means intended to the target population. For example, an intervention may be effective when delivered by highly selected and trained staff and extensive resources but in routine clinical practice this may not be achievable.

Effectiveness & Cost-Effectiveness

Effectiveness refers to the effect size of the intervention in relation to desired objectives in a real world context. It is distinct from efficacy which refers to the effect size of the intervention when delivered under optimal conditions in comparative evaluations.

Cost-effectiveness refers to the ratio of effect (in a way that has to be defined, and taking account of differences in timescale between intervention delivery and intervention effect) to cost. If two interventions are equally effective then clearly the most cost-effective should be chosen. If one is more effective but less cost-effective than another, other issues such as affordability, come to the forefront of the decision making process.

Acceptability

Acceptability refers to the extent to which an intervention is judged to be appropriate by relevant stakeholders (public, professional and political). Acceptability may differ for different stakeholders. For example, the general public may favour an intervention that restricts marketing of alcohol or tobacco but politicians considering legislation on this may take a different view. Interventions that appear to limit agency on the part of the target group are often only considered acceptable for more serious problems.

Side-effect/Safety

An intervention may be effective and practicable, but have unwanted side-effects or unintended consequences. These need to be considered when deciding whether or not to proceed.

Equity

An important consideration is the extent to which an intervention may reduce or increase the disparities in standard of living, wellbeing or health between different sectors of society.

A.2.3 7 Situational Context questions from the BCW framework

1. "Where does the behaviour occur?"
2. "Who is involved in performing the behaviour?"
3. "What do they need to do differently to achieve the desired behaviour?"
4. "Where will they do it?"
5. "Who needs to perform the behaviour?"
6. "With whom do they need to do it?"
7. "When do they need to do it?"

A3. AI & GPT

A.3.1 History of AI development

The first works on the topic of Machine Learning (ML) can be traced back to 1943 "A logical calculus of the ideas immanent in nervous activity" by Walter Pitt and Warren McCulloch. Ever since the field has been developing. Important early publications such as the book "The organisation of behavior" (1949) by Donald Hebb, Alan Turing published work in 1950, "Computing Machinery and Intelligence", with its famous Turing Test, and in 1952 the Hodgekin-Huxley model. Which at a conference in 1956 sparked the concept of artificial intelligence. The first Neural Network was built by Frank Rosenblatt in 1957 "the perceptron", which simulated the thought processes of the human brain. And Henry J. Kelly (1960) published "Gradient theory of optimal flight paths" which laid the foundations of the continuous Back Propagation model (only useful after 1985). Progress was slow, and the next significant step was in 1967, when the "Nearest Neighbour" algorithm was proposed by Thomas Cover.

The lack of significant progress due to limitations in lacking memory and processing power caused a cut in government research funding in the 1970s and led to the 'The first AI winter'. Some however continued researching the topic without funding and in 1980, Kunihiko Fukushima designed the Neocognitron based on convolutional neural networks, a hierarchical multi-layered design. Many concepts of Neocognitron continue to be used.

In the 1980s the research was resumed through the US and UK funded Japanese "fifth generation" computer project. And breakthroughs started to emerge from universities such as the 'Stanford Cart' in 1979 and the publication of Explanation Based Learning (EBL) by Gerald Dejong in 1981. DARPA (Defense Advanced Research Projects Agency) however decided to cut funds again between 1987 and 1993, which is referred to as 'the second AI winter'. Some research however continued and in 1989, Yann Le Cun introduced the first demonstration of back propagation combined with convolutional neural networks at Bell Labs, it allowed for reading out handwritten numbers from checks.

The concept of boosting, by Robert Shapire (1990) from his publication "The Strength of Weak Learnability" and increased memory and processing power. Led to a shift from a knowledge-driven to a data-driven approach. Causing Natural Language Processing (NLP) to evolve from decision trees into statistical models. These Big Data programs gradually shifted from bots to virtual assistants and chatbots.

In 1995, Dana Cortes and Vladimir Vapnik developed the support vector machine (a system for mapping and recognizing similar data). Around the 2000s the 'Vanishing Gradient Problem' appeared which was caused by large areas of input being condensed into a small range of output, and caused action functions to vanish in the data gradient. Two solutions were developed for this problem, layer-by-layer pre-training and the development of long short-term memory (LSTM). LSTM (long short-term memory) for recurrent neural networks was developed in 1997, by Sepp Hochreiter and Juergen Schmidhuber, the same year that IBM's Deep Blue system beat the world champion at chess. The introduction of GPU's by Nvidia (1999) increased the computational power a 1000 times of a time span of 10 years.

Deep Learning (DL) was coined by Geoffrey Hinton in 2006 to explain new algorithms that let computers "see" and distinguish objects and text in images and video. Fei-Fei Li, a Stanford professor in AI launched ImageNet in 2009, which contained 14 million labeled images suitable for 'supervised learning' pre-training neural nets. In 2010, Microsoft Kinect was able to track 20 human features at a rate of 30 times per second.

Increased GPU speed allowed for training convolutional neural networks without Layer-by-layer pre-training. An early example of this 'unsupervised learning' is the deep neural network Google Brain project 'the cat experiment', developed in 2012, which used unlabelled images to learn to discover and categorize objects.

Siri turned into a highly popular digital virtual assistant (2011). In 2014 Facebook developed DeepFace, which was able to recognize and verify faces similar to humans.

In 2014 Generative Adversarial Neural Network (GAN) was introduced by Ian Goodfellow. In this concept two neural networks play against each other in a game and is a mechanism which is used to optimize output.

In 2015 Microsoft created the Distributed Machine Learning Toolkit, Amazon launched its own machine learning platform. That same year 3,000 AI and Robotics researchers, endorsed by Stephen Hawking, Elon Musk (among others) signed a letter to warn for the danger of autonomous weapons, which select and engage without human intervention.

OpenAI

In 2015 also OpenAI was founded with an initial pledge of 1 billion US\$, an American AI research laboratory consisting of a non-profit OpenAI Incorporated (OpenAI Inc.) and its for-profit subsidiary corporation OpenAI Limited Partnership (OpenAI LP). In 2019 Microsoft invested 1 billion US\$ in OpenAI LP, and in 2023 a multiyear investment reported to be 10 billion US\$. Due to the humanitarian mission of the OpenAI group it was able to acquire top level AI researcher.

In April 2016 OpenAI released 'OpenAI Gym' a public beta for reinforcement learning. In December 2016 OpenAI released 'Universe' for training general intelligence across the games, websites and other applications. Between 2017 and 2020 OpenAI created several research tools such as RoboSumo, OpenAI Five, Gym Retro, Debate Game and Dactyl.

In 2020 a multi-purpose API was announced which was "for accessing new AI models developed by OpenAI".

A3.2 Summary of GPT

GPT stands for 'Generative Pre-trained Transformer' and are autoregressive, neural network, machine learning models trained on data from the internet using Natural Language Processing (NLP). NLP is a field of computer science which aims to give computers the ability to understand human language by combining; Computational Linguistics (CL) (Church & Liberman, 2021), the science and engineering of understanding written and spoken language from a computational perspective; Machine Learning (ML), in which the human way of learning is imitated by computers; and Deep Learning (DL) (Janiesch et al. 2021), the science of human brain inspired computational processes; inside of a large language model (LLM), which are trained on vast amounts of data (GPT-3 was trained on 175 billion parameters), allowing it to learn patterns and relationships which it uses to create predictions and groupings.

LLMs like GPT are pre-training via 'unsupervised learning' on vast data sets (a significant part of the internet), according to a fine-tuning process. GPT then stores the information and its underlying relations in the neural network of the LLM, which accesses to predict which output is most likely to follow from the input (Petroni et al., 2019). These LLMs are autoregressive, meaning that they are capable of updating their neural network according to the input they are provided with.

A3.3 Prompt engineering

A3.3.1 Task specification

The input that is provided to an LLM is called a prompt, the science of constructing a prompt to get desirable output is called prompt engineering. An important difference in prompt designs are the so-called, zero-shot, one-shot and few-shot prompts (Brown et al., 2020).

The few-shot prompts or '*task specification by demonstration*' originates from traditional ML science. Researchers provide the LLM with a set of examples, and the LLM provides an output which is intended to be a continuation of the examples on the same semantic levels. These training examples can be labeled to train the LLM on the different nature of the provided examples, for instance by labeling examples pro-bias or anti-bias, to reduce bias in the output of the LLM (Hendrycks et al., 2021b).

Zero-shot prompts or '*direct task specification*' are used by many. It uses a natural language construct to give the LLM a direct task (e.g. after a direct question, the LLM will predict the most likely answer to this question). It requires a signifier, which is a textual pattern that initiates the intended behavior. Direct task specifications are compact, but draw from infinite implicit examples from the neural network, making them powerful when performing more open tasks.

Another form of zero-shot prompting which is closely related to direct task specification is *specification by memetic proxy*. GPT-3 has shown a good understanding of analogies (Czinczoll, 2022). And as such, GPT-3 is able to simulate public figures and access cultural understanding beyond the capability of humans (Branwen, 2020). Analogy is a common strategy in human communication, and can be utilized in engineering prompts. An example of this is prompting the LLM to act from the persona of a dietician, the LLM will locate its task within the neural network based on its understanding of how a dietician acts. And will shape its responses based on complex biases and assumptions about the context of a question. (White et al., 2023)

Brown et al. (2020) from the OpenAI University provides evidence that few-shot prompts can produce better results than zero- and one-shot prompts. Explaining that supplying the LLM with a set of examples of the output allows the LLM to learn the task at runtime, and that as such, more examples result in a higher performance.

However, Reynolds et al. (2021) shows that 0-shot prompts can significantly outperform few-shot prompts. They reason that instead of instructing the LLM to learn, the examples are supporting task allocation for an already learned task. Reynolds et al. (2021) do not disregard the effectiveness of few-shot prompts, and suggest *that task specification by demonstration* is a way to prompt an LLM to perform behaviour for "which it is less obvious to construct a direct signifier".

GPT model settings

The GPT technology offers a 0-1 temperature scale which can be used to influence the behaviour of the model. It influences the probability tolerance, or Markov assumption (Markov, 1945), the model makes. For instance, a low temperature setting of 0,1 will only result in highly probability outputs, in turn a high temperature setting of 0,9 results will results in outputs with a lower probability. The temperature scale not similar but shows correlation to the 'cleverness' and 'creativity' of the model (Si et al, 2022). For this study the default setting of 0,7 is used, since this setting allows for enough variation in the output, but refrains from highly improbable results. An additional argument for this setting, is that it is the default setting and as such is what most people use when they interact with the GPT model.

A3.3.3 Prompt Patterns

Reynolds et al. (2021) further state that an issue with LLMs is that they do not respond to a prompt *like a person would*, but like *any person could*. As such **the prompt should not only be consistent with the desired continuation, but also inconsistent with any undesired continuation of the prompt**. Therefore approaching the instructions in a prompt as programming can support in directing and refining its capabilities (Liu et. al, 2023). In software the use of patterns provides us with reusable solutions for recurring problems (Gamma et. al, 1995). And the same can be true for the patterns of a prompt. White et al. (2023) provides us with a catalogue of essential domain independent prompt patterns ([Appendix A5.2.1](#)).

A3.4 Reliability of LLMs

Since GPT-3 and GPT-4 have not reached consistent 'ground truth' (outputting correct or 'true' answers to all prompts), it is highly important to consider its reliability. Si et al. (2022) assess reliability via a framework inspired by a survey of unsolved problems in ML safety (Hendrycks et al., 2021b). Namely; *Withstanding hazards* or 'generalizability' about how well an LLM responds to unseen domains or other challenging examples; *Identifying hazards* or 'uncertainty calibration' about the confidence of the LLM about its output predictions; And *steering ML systems and reducing deployment hazards* which can be divided into 'social bias and fairness' and 'factuality'.

A3.4.1 Generalizability

Generalizability is highly important for multifunctional deployment of an LLM, as well as for more open tasks which require the LLM to deal with high ambiguity. Its capability in generalizability is enhanced when a wide variety of examples is provided (Si et al., 2022). This however holds less relevance for training an LLM for a specific task such as the generation of motivational text messages.

A3.4.2 Confidence calibration

Confidence calibration is, since language models are imperfect, necessary in order to safely trust LLMs, especially in high stakes situations. Its confidence can be determined by calculating the probability of which word can follow a particular sentence, called a 'Markov assumption' (Markov, 1945). The OpenAI playground interface ([source](#)) which gives access to the various LLMs from OpenAI, offers the *hyperparameter 'temperature' scale*, ranging from 0 to 1 in order to control the probability of the output. OpenAI warns against misinterpreting the temperature setting as "'cleverness' or 'creativity' controls". But Si et al. (2022) do provide some evidence about a correlation between the confidence scores and factual accuracy. The popular publicly accessible ChatGPT interface has a standard temperature setting of 0,7 ([source](#)).

Low temperatures result in outputs with higher Markov probabilities, resulting in conservative, deterministic results which are low in variability and that are good at answering questions in which only one answer is correct. Conversely a high temperature results in output with a lower Markov probability and results in increased randomness and more unpredictable outputs. (OpenAI prompt design guide, z.d.). The temperature setting is thus dependent on the use case of the prompt. For the generation of motivational text messages a high variability is desirable, while it has to stay within the scope covering the format constraints, target behaviour and behaviour change technique.

A3.4.3 Social bias and fairness

Social bias and fairness are a dominant part of the public discussion about emerging AI technology. Especially since 'unsupervised learning' on a large part of the internet is partial to the success of OpenAI's GPT models. They consequently can be prone to cause bias and unfair output results due to lacking, outdated or wrongly interpreted data (Chan, 2022). If bias models are deployed into real situations they can cause serious harm, especially to minority groups (Cao et al., 2022).

Si et al. (2022) significantly reduced social bias and fairness in the LLMs output by feeding it several, balanced examples with pro-bias and anti-bias labels. For an optimal performance (lowest bias gap) the examples had to be shuffled into a random order. Additionally it was found that a proper distribution of demographic biases in the examples was essential and that a disbalance increased the bias gap. Lastly, the addition of a natural language instruction in the prompt (ref. 1) about how to treat the output regarding bias also showed a significant reduction of the bias gap.

Ref 1. "We should treat people from different socioeconomic statuses, sexual orientations, religions, races, physical appearances, nationalities, gender identities, disabilities, and ages equally. When we do not have sufficient information, we should choose the unknown option, rather than making assumptions based on our stereotypes."

A3.4.4 Factuality

That the *Factuality* of LLM outputs are far from trustworthy, and that its inaccuracy can have serious consequences is illustrated by the \$100bn dollar factual error made by Google's AI Bard in a prerelease promotion video (BBC News - Sherman, 2023). But also Microsoft's GPT integration into Bing is far from flawless (CNBC, Lewing, 2023).

The knowledge used to train an LLM can be faulty or outdated and the relations within its neural network might be misinterpreted. However, more often the size of the LLM causes factual inaccuracies, due to misunderstanding definitions and niche terminology in specialist tasks (Abildgaard, 2023). One solution is to develop specialized models trained on curated data niche tasks, which requires highly specialized skills and expertise, which are rare and in high demand.

Since LLMs like GPT are autoregressive models, another solution can be found in providing the LLM with correct contextual information within the input prompt. Si et al. (2022) show that this approach consistently improves its performance in *factuality*.

Additionally Si et al. (2022) show that when multiple steps of reasoning are required, a LLM might take a wrong turn in reasoning. The risk of this happening can be partially mitigated by approaching this with a Chain-of-Thought (CoT) prompt, CoT means that an LLM decomposes its task into steps, each time building upon the output from its previous task, much like a human would use a scratch pad to solve a complex task. CoT allows us to elicit the reasoning process a LLM uses to solve a task (Wei et al., 2022b).

A4. Experimental Generative Prompts

A4.0 Process overview

1 - For the **first iteration** two direct task 0-shot prompts were experimented with in order to assess GPTs capability to create MTMs, resulting in properly formatted and formulated but generic MTMs ([Appendix A4.1](#)).

2 - The **second iteration** attempted to mimic the BCW process in a few-shot prompt with examples from the MacPherson study. The prompt specified the; Target Behaviour (TB), Behaviour Change Technique (BCT) and answered the seven situational context questions from the BCW ([Appendix A2.3](#)). The prompts applied a 'lifestyle coach' analogy and utilised a Q&A structure (OpenAI API. (z.d.-c) to frame the situational context. The TB and the seven context question answers were specified by tasking GPT to generate many target behaviours. After selecting one based on personal preference it was tasked to elaborate on it and to answer each of the context questions for that TB. The output was then curated and summarised before being integrated into the prompt. The resulting MTMs this prompt generated contained specifics regarding the behaviour change, but appeared to be generally applicable ([Appendix A4.2](#)).

Factuality: [Appendix 3.4.4](#)

3 - For the **third iteration** these elaborate prompts were decomposed into modules and mapped on eight functionality groups in a morphological chart (Zwicky, 1940). After evaluating the project scope and research method three groups were discontinued.

- 'Personalization' functions such as a *personality profile*, *personal goals* and *progress* are promising elements in which GPT models might excel. However, their inclusion would compromise the comparability of the baseline MTMs.
- 'Referencing' the *BCW theory and process*, *specifying the intervention function* and *explaining how a task is part of the BCW framework* were excluded, as the GPT model had too little understanding integrated into its neural network, rendering it unable to recall and contextually understand.
- 'Context' related functions such as a *general personas*, *generic user scenarios* and the *7 situational context questions*, were excluded regardless of their successful output in the prior iteration. Since we want to understand how well the GPT model functions, it is better to minimise the prompt and have it rely on its own neural network, compared to providing it with a detailed framework of the context it should consider. Which also allows for better replicability of the study.

For the remaining functions Minimum Viable Prompts (MVPs) were created, to test their individual performance ([Appendix A4.3](#)).

4 - During the **fourth iteration** three minimalistic prompts were created to test different task specifications. Each prompt was based on the same minimal system message, naming the TB, BCT and format constraint, additionally a 'bias prevention' text ([Appendix A3.4.3](#)) was added to minimise the risk of unethical outputs, as they were tested with students ([Appendix A4.4.1](#)). The three prompts consisted of two 0-shot prompts, of which the first used a direct task specification, the second a 'Dietician' analogy, and the third, a few-shot prompt, used six MTMs from the MacPherson study of the same BCT ([Appendix A4.4.2](#)). These were tested during a student fair using a physical card sorting method on a 7 point likert scale ([Appendix A4.4.3](#)).

5 - Based on the peer feedback from the pilot study a **fifth iteration** revolved around reconsidering the research method and the corresponding prompt. To ensure that the results are comparable, a combination of task specifications is required. For this a task specification with dietician analogy will be combined with a specification by demonstration in the form of a 1-shot prompt.

The reason for this is that each validated message is designed with a specific behaviour change state in mind. As such in order to ensure that GPT is prompted to generate comparable messages, only one message should be provided as an example to provide some contextual guidance.

To ensure we account for the variety in the messages, three prompts with one message will be tested. For each prompt 5 messages will be generated and will be referred to as parent (Validated message) and child (Generated message). As such three surveys will be conducted each containing one parent MTM and its corresponding child MTMs. (Appendix A4.5).

6 - The last, **sixth iteration** considered *quantitative* (Appendix A5.1) and *qualitative* (Appendix A5.2) reflective and evaluative prompt elements. After several experiments prompting GPT to grade its own messages on the APEASE criteria, the quantitative route was discontinued. They are, however promising, too inconsistent to be usable, and require further development before implementation. The qualitative evaluation and reflection modules borrowed from White et al. (2023) however proved insightful (Appendix A5.2.3) and were added together with an extra format prompt to consistently structure the output into an easy to manage structure (Appendix A5.2.2).

A4.1 Iteration 1

A4.1.1 Test 1

The first prompt attempt was minimalistic, only containing a task specification (Reynolds et al, 2021 & Brown et al.,2020) and a format constraint. It showed that the GPT technology is capable of generating properly formulated, somewhat comparable MTMs, to the baseline MTMs. The generated MTMs were however highly unspecific, due to a lacking contextual framing.

INPUT

"Write 10 motivational text messages of maximum 160 characters about goal setting."

OUTPUT

"Don't be afraid to dream big. With hard work, determination, and a solid plan, you can achieve anything you set your mind to."

A4.1.2 Test 2

The second prompt was written based on the initial study scope. Its composition combined a task specification, format constraints, a situational context, overall intention, references the BCW and names the BCT. Due to the situational context, the messages were more personalised to a specific use case than the baseline. Next to this it was found that the messages were not specific in guiding the target behaviour.

INPUT

"Write 10 motivational text messages of maximum 160 characters for someone that has just finished year long intervention, with the intention of changing their behaviour in nutritional intake and physical activity. The messages will be part of a bigger message bank, all designed according to the Behaviour Change Wheel Framework. Write this set of messages for the Behaviour Change Technique: 1.1. Goal setting (behaviour)."

OUTPUT

"Setting goals can be challenging, but you've already shown you're capable of making changes. Let's set some new goals and continue your progress."

A4.2 Iteration 2

A4.2.1 Test 3

In the third prompt a persona based task specification was combined with the eight main questions from the BCW theory framed within a situational context in a Q&A structure (OpenAI API. (z.d.-c) . The answers to these questions were initially formulated top of mind. Additionally the prompt references the same intervention functions from the baseline study, and next to the BCT name also the BCT description. The output MTM was too personalised, and as such more generally applicable. But offered more guidance for achieving the target behaviour.

INPUT

“You are a lifestyle coach, you are coaching according to the Behaviour Change Wheel framework. After an intake of a client you have specified the target behaviour via the following eight questions:

Q. What is the target behaviour?

A. The target behaviour is 'Improve nutritional intake'

Q. Where does the behaviour occur?

A. The behaviour occurs 'Anywhere where we eat or buy food'

Q. Who is involved in performing the behaviour?

A. The ones involved in performing the behaviour is 'Everyone that lives together in the same house hold'

Q. What do they need to do differently to achieve the desired behaviour?

A. The behaviour they need to do differently to achieve the desired behaviour is 'Adhere constantly to a heart healthy diet'

Q. Where will they do it?

A. They will do it 'At home, in the kitchen, at the grocery store, at the canteen and in restaurants or take-out places'

Q. Who needs to perform the behaviour?

A. The one that will do it are 'Healthy adults that struggle with their body weight' Q. With whom do they need to do it?

A. They will do it with 'Friends and family'

Q. When do they need to do it?

A. They will do it during 'Breakfast, Lunch or Dinner time, or when buying food'

Based on the framework you selected the following intervention functions:

- *Environmental*
- *Restructuring*
- *Incentivization*
- *Enablement*
- *Education*
- *Training*
- *Persuasion*

You are going to write a motivational message for the client, and have selected the behaviour change technique 1.3, Goal setting (outcome) for this.

This behaviour change technique description is to 'Set or agree a goal defined in terms of the behaviour to be achieved' the goal needs to define a specific context, frequency, duration or intensity. Write 10 motivational text messages of maximum 160 characters per message.

OUTPUT

"We can improve your nutritional intake by setting a goal to plan your meals in advance and make a shopping list to ensure you have healthy options on hand."

A4.2.2 Test 4

For the fourth prompt the same format as the third prompt was used. The difference was the addition of some validated example messages, changing the prompt from a zero-shot to a few-shot prompt. The behaviour change technique was changed because it had more validated examples available. The differences between the prompt outputs are harder to see from this point on.

INPUT

"[similar to prompt three] + You are going to write a motivational message for the client, and have selected the behaviour change technique **Social support (unspecified) (3.1)** for this.

This behaviour change technique description is to '**Advise on, arrange or provide social support (e.g. from friends, relatives, colleagues, 'buddies' or staff) or non-contingent praise or reward for performance of the behaviour. It includes encouragement and counselling, but only when it is directed at the behaviour. Note: attending a group class and/ or mention of follow-up' does not necessarily apply this BCT, support must be explicitly mentioned;**'

Write 10 motivational text messages of maximum 160 characters per message.

Here are some example messages:

1. Sometimes we need the support of others to help us stick to our goals. Think about who in your life can help you stick to your exercise plan.
2. Having people who support us is important! Think about who in your life can help you stick to your food goals.
3. Maintaining healthy behavior can be challenging. Many people find it helpful to share this journey with a close friend.
4. Take a moment to think about all the hard work you've put in. You are becoming a healthier you!"

OUTPUT

"Remember, you don't have to go through this alone. Lean on your loved ones for motivation and encouragement."

A4.2.3 Test 5

The fifth prompt had no structural changes, but the answers to the prompt questions were finetuned and enriched. For this GPT was prompted in a way that it mimics the first steps of the BCW approach.

INPUT

“You are a lifestyle coach, you are coaching according to the Behaviour Change Wheel framework. After an intake of a client you have specified the target behaviour via the following eight questions:

Q. What is the target behaviour?

A. The target behaviour is 'Improve dietary habits: Encouraging people to eat a healthy and balanced diet that is low in calories, saturated and trans fats, and sugar, and high in fruits, vegetables, whole grains, and lean protein.'

Q. Where does the behaviour occur?

A. 'The behavior of improving dietary habits to prevent obesity occurs primarily in the individual's daily life and environment. This includes their choices around food selection, meal preparation, and consumption patterns. Additionally, the behavior of improving dietary habits may also occur in a variety of social and environmental settings, such as in schools, workplaces, and communities.'

Q. Who is involved in performing the behaviour?

A. The ones involved in performing the behaviour are 'The primary person involved in performing the behavior is the individual seeking to improve their dietary habits to prevent obesity. Additionally family members or friends who may influence the individual's food choices or meal preparation.'

Q. What do they need to do differently to achieve the desired behaviour?

A. The behaviour they need to do differently to achieve the desired behaviour is 'Make healthier food choices; Limit high-calorie, high-fat, and high-sugar foods; Pay attention to portion sizes; Plan and prepare meals; Eat mindfully; Seek support; Stay motivated'

Q. Where will they do it? A. They will do it 'At home; At work; In restaurants and food establishments' Q. Who needs to perform the behaviour?

A. 'The behavior of improving dietary habits to prevent obesity can be performed by anyone who wishes to make positive changes to their diet and prevent obesity. While some individuals may be at higher risk for obesity than others, such as those with a family history of obesity or certain medical conditions, anyone can benefit from making healthier dietary choices. Improving dietary habits to prevent obesity is especially important for healthy adults who are not currently overweight, but who want to maintain a healthy weight and prevent obesity from developing. This includes individuals who may have a sedentary lifestyle, unhealthy eating habits, or who may be at risk for chronic diseases like diabetes or heart disease.'

Q. With whom do they need to do it?

A. They will do it with 'Family, friends and Co-workers'

Q. When do they need to do it?

A. 'Improving dietary habits to prevent obesity is a behavior that should be practiced consistently over time in order to achieve and maintain long-term benefits. They need to do it Daily; during Mealtime; during Grocery shopping; and during Dining out: even during Special occasions.'

You are going to write a motivational message for the client, and have selected the behaviour change technique Social support (unspecified) (3.1) for this. This behaviour change technique description is to 'Advise on, arrange or provide social support (e.g. from friends, relatives, colleagues, 'buddies' or staff) or non-contingent praise or reward for performance of the behaviour. It includes encouragement and counselling, but only when it is directed at the behaviour. Note: attending a group class and/ or mention of follow-up' does not necessarily apply this BCT, support must be explicitly mentioned;'

Write 10 motivational text messages of maximum 160 characters per message. Here are some example messages:

1. Sometimes we need the support of others to help us stick to our goals. Think about who in your life can help you stick to your exercise plan.
2. Having people who support us is important! Think about who in your life can help you stick to your food goals.
3. Maintaining healthy behavior can be challenging. Many people find it helpful to share this journey with a close friend.
4. Take a moment to think about all the hard work you've put in. You are becoming a healthier you!

OUTPUT

“Surround yourself with people who believe in your ability to succeed. Their support can make all the difference in achieving your healthy eating goals.”

A4.3 Iteration 3

To have a prompt generate desirable MTMs, it has to be composed out of several text components. By systematically segmenting the experimental prompts into parts, we can distinguish components and plot them onto a morphological chart (Zwicky, 1940) (Appendix 4.3.1).

The components in the Personalised and Grounded groups were discontinued. The Personalised components are very interesting, but are outside of the scope of this study. The Grounded components might become relevant in future versions of the GPT technology, however the GPT 3.5 version which was used for testing these modules had no explicit understanding of the referenced material, defying the purpose of including it into the prompt.

Next to this the 'context' elements from situatedness were excluded. The reasoning for this is that any contextual framing will strongly influence the outcome. An increased contextual granularity will cause the output of the GPT technology to be directed towards these contextual elements. Which shows on the one hand that the output of GPT technology can be controlled, to a certain extent. But it also will distort the results of this study due to increased complexity.

A4.3.1 Morphological Chart

Group	Function	Options			
FIT	Format	<i>Character limit</i>	<i>Write (X) nr. of MTMs</i>		
Targeted	Strategy	<i>Name BCT</i>	<i>Specify BCT</i>		
	Example	<i>One validated example</i>	<i>Few validated examples</i>	<i>Labelled examples</i>	
Situatedness	Persona (analogy)	<i>Dietician</i>	<i>Lifestyle coach</i>	<i>Behaviour change expert</i>	<i>System description</i>
	Context	<i>General persona</i>	<i>General User Scenario</i>	<i>Naming the 8 BCW Q&A's</i>	<i>Specify the 8 BCW Q&A's</i>
	Goal	<i>Name Target Behaviour</i>	<i>Short Target Behaviour description</i>	<i>Specify Target Behaviour</i>	
Personalised	User details	<i>Personality profile</i>	<i>Personal goals</i>	<i>Progress</i>	
Grounded	Reference	<i>Specify Intervention functions</i>	<i>Part of bigger whole</i>	<i>BCW theory and process.</i>	

Table x. Morphological chart: generative prompt design

A4.3.2 Testing the behavior of separate modules

Testing separate modules [revise]

To understand how each separate component behaves, 12 minimal prompts were created containing only the FIT components and one of the Targeted or Situatedness components. For the BCT and Target Behaviour several granularities were explored, to understand how detailed their specification should be.

In one prompt the format constraint was excluded to understand how well GPT would commit to the provided example format structure ([Appendix 4.3.2](#)).

Since the test focuses on nutritional intake, and the study is performed with dieticians, the Dietician persona unsurprisingly performed best on the Roleplay prompt testing.

For the Directed Situated prompt only naming the target behaviour was selected.

Specifying the target behaviour however resulted in specific results, which was excluded for the same reason the contextual elements were excluded. Adding (relatively arbitrary) specific contextual elements into a prompt strongly influences the results, increasing the complexity of the results while decreasing the comparability of the prompt.

For the Behaviour change specification both naming and specification were selected. The specification in this context was considered good because the GPT technology needs constraints to understand which task it has to perform. However, contrary to the Target Behaviour, the BCT is not guided by specific contextual elements, it is rather providing a solution direction, which is required to create comparable results.

Providing GPT with examples did result in more comparable messages. Without them it struggles to understand with the tone of voice. It might for instance start using emojis or hashtags which are probably common practice in the training data of GPT, but are not appropriate for comparison to the messages.

The task specification is an essential part of the prompt, but so is the format constraint, as GPT showed inconsistent coherence to the 160 characters rule when no format constraint and only examples were provided.

SITUATEDNESS				TARGETED							
ROLEPLAY PROMPT TESTING				DIRECTED PROMPT					UNDIRECTED PROMPT		
1	2	3	4	5	6	7	8	9	10	11	12
FIT	FIT	FIT	FIT	FIT	FIT	FIT	FIT	FIT	FIT	FIT	FIT*
<i>Dietician</i>	<i>Lifestyle coach</i>	<i>Behaviour change expert</i>	<i>System description</i>	<i>Name Target Behaviour</i>	<i>Short Target Behaviour description</i>	<i>Specify Target Behaviour</i>	<i>Name BCT</i>	<i>Specify BCT</i>		<i>Few validated examples</i>	<i>Few validated examples</i>

A4.4 Iteration 4

A4.4.1 Fundamental Prompt

Every prompt requires some form of task specification, Reynolds et al. (2021) suggest three different forms, namely; specification by memetic proxy, direct task specification and specification by demonstration (Appendix A4.4.1).

bias prevention	:“We should treat people from different socioeconomic statuses, sexual orientations, religions, races, physical appearances, nationalities, gender identities, disabilities, and ages equally. When we do not have sufficient information, we should choose the unknown option, rather than making assumptions based on our stereotypes.”
target behaviour	:Improving dietary habits
Behaviour change technique	:Social Support
format constraint	:Use a maximum of 160 characters.

A4.4.2 Prompt Task Specifications

specification by memetic proxy	Act as a dietician write one motivational text message that a dietician would create.
direct task specification	Write one motivational text messages
specification by demonstration	<ul style="list-style-type: none"> • "Sometimes we need the support of others to help us stick to our goals. Think about who in your life can help you stick to your exercise plan." • "You've shown yourself that you can integrate exercise into your daily routine. Keep up the great work!" • "Having people who support us is important! Think about who in your life can help you stick to your food goals." • "Maintaining healthy behavior can be challenging. Many people find it helpful to share this journey with a close friend." • "Take a moment to think about all the hard work you've put in. You are becoming a healthier you!" • "Exercising with a buddy can give you motivation and support - and can be a nice distraction!" <p>[...]</p>

A4.4.3 Test set-up

In order to determine which one was more fitting, six generated MTMs for each of the task specifications were compared with six validated handcrafted MTMs during an early pilot study at university fair to receive peer feedback.

Each MTM was printed on a separate card, bystanders were then asked to choose one of the 6 APEASE criteria and perform a closed card sort, which was comparable to a 7-point Likert scale survey. This execution served as a pilot study to test both study set-up and prompt design. **The exercise was performed with several students (4) and feedback was received from an Behaviour change Expert and from an expert on statistics in social studies.**



A4.5 Iteration 5

From the pilot several insights were drawn, in essence they relate back to the method more than the design. However, to ensure that the results are comparable, a combination of task specifications is required. For this a memetic task specification will be combined with a specification by demonstration in the form of a one-shot prompt.

The reason for this is that each validated message is designed with a specific behaviour change state in mind. As such in order to ensure that GPT is prompted to generate comparable messages, only one message should be provided as an example to provide some contextual guidance.

To ensure we account for the variety in the messages, three prompts with one message will be tested. For each prompt 5 messages will be generated and will be referred to as parent (Validated message) and child (Generated message). As such three surveys will be conducted each containing one parent MTM and its corresponding child MTMs.

4.5.1 Fine Tuned Generative Prompt Design

Prompt strategy

For the eventual design of the prompt a one-shot approach was chosen, because after evaluation its outputs can be labeled and used as examples for a new prompt to improve and fine-tune its results. For the prompt the one baseline message was provided as examples. The “quotation marks” enable the model to differentiate between the examples. The [...] punctuation has shown to be an effective way to make the model understand that it has to generate its content as a continuation of the provided examples.

The prompt was enhanced by providing it with a 'system message' containing several elements to clarify the intention and provide constraints:

1. A natural language bias prevention frame (Si et al., 2022)
2. A 'direct task specification with persona' (Reynolds et al., 2021) & (White et al., 2023)
3. A format constraint (MacPherson et al., 2021c)
4. The name of the target behaviour (MacPherson et al., 2021c)
5. The name of the behaviour change technique and a short description (Michie et al., 2014)
6. Specification by baseline message

<i>nr.</i>	<i>Pattern name</i>	<i>Prompt text</i>
1	<i>Bias prevention</i>	"We should treat people from different socioeconomic statuses, sexual orientations, religions, races, physical appearances, nationalities, gender identities, disabilities, and ages equally. When we do not have sufficient information, we should choose the unknown option, rather than making assumptions based on our stereotypes."
2	<i>direct task specification with persona</i>	Act as a dietician write one motivational text message that a dietician would create.
3	<i>Format constraints</i>	Use a maximum of 160 characters.
4	<i>Target behaviour</i>	I would like to achieve the improvement of dietary habits
5	<i>BTC</i>	According to the behaviour change technique: 'Social support'
6	<i>BTC description</i>	This behaviour change technique is described as follows: 'Advise on, arrange or provide social support (e.g. from friends, relatives, colleagues, buddies' or staff) or non-contingent praise or reward for performance of the behaviour. It includes encouragement and counselling, but only when it is directed at the behaviour. Note: attending a group class and/ or mention of follow-up' does not necessarily apply this BCT, support must be explicitly mentioned;'
0	<i>specification by baseline messages</i>	Comparable to: "Sometimes we need the support of others to help us stick to our goals. Think about who in your life can help you stick to your exercise plan." [...]

A5. Evaluative Prompt design - iteration 6

In oncology narrow-task AI technology shows high potential as it is being used to evaluate patient data to support diagnoses (Kann et al. 2021). Design thinking is a reflective practice (Schön, 1983), as such an AI supporting this practice could also yield high potential for the design practice of this study.

Would a LLM trained on vast data sets be able to evaluate MTMs on their Intervention Effectiveness or its underlying components? Regrettably, up to this moment, only to a certain extent.

A5.1 Experiment examples quantitative prompt (Review)

Several experiments were performed (Appendix A5.1), but many had to be disregarded after realising that ChatGPT had a word limit of roughly 3000 words. Next to this it was found that the experiment in which GPT was tasked to assign a value via a given scale on an MTM was too inconsistent to use. More research is required on this, because the capability of GPT technology to evaluate large quantities of texts on specific criteria would not only transform the field of design, but would transform our interaction with GPT and its perceived reliability.

A few findings were worth sharing for future attempts.

- GPT appears to have many different understandings for one word, especially more latent variables. A thorough description of these criteria is therefore an essential addition.
- GPT lends itself as an efficient tool to generate these descriptions for you, as it is capable of describing its many understanding of a word for you. By curating its understanding it seems that we can narrow down GPTs word definition.
- Another approach would be to ask GPT to decompose a subject into elements, parts or factors, allowing us to elicit its underlying understanding.
- GPT seems to become more consistent if it is tasked to use rubrics for its evaluations. GPT itself can also be used to write these rubrics for us.
- A low temperature setting will ensure that the system adheres more strongly to the provided rubrics.

Since for this iterative cycle between 20 and 30 experiments with long complex prompts were conducted, only a few are included in this appendix to convey the general approach to the reader. The conclusion was drawn that a quantitative self-assessment mechanism was not achievable for this particular study. More focused research is required to assess how and if this will be feasible for future projects.

A5.1.1 Experiment 1:

GPT was provided with the definition of the six APEASE criteria from the BCW theory, ten MTMs, and the task to 'Assign a 1-10 grade to the criteria per message'. The results were a string of numbers of which it was impossible to tell how reliable they were.

A5.1.1. 2 Experiment 2:

GPT was tasked to rephrase the criteria Affordability and its description into rubrics for each of the 10 grades.

1. *Not acceptable: The intervention is not affordable, and it cannot be delivered to all those who would benefit from it.*
2. *The intervention is minimally acceptable, but it may be difficult to deliver to all those who would benefit from it due to budget constraints.*
3. *The intervention is somewhat acceptable, but it may require some budgetary adjustments to make it accessible to all those who would benefit from it.*
4. *The intervention is acceptable, but it may require some budgetary adjustments to make it accessible to all those who would benefit from it.*
5. *The intervention is moderately acceptable, and it can be delivered within a reasonable budget to most of those who would benefit from it.*
6. *The intervention is reasonably acceptable, and it can be delivered within a reasonable budget to most of those who would benefit from it.*
7. *The intervention is quite acceptable, and it can be delivered within a reasonable budget to almost all those who would benefit from it.*
8. *The intervention is highly acceptable, and it can be delivered within a reasonable budget to almost all those who would benefit from it.*
9. *The intervention is extremely acceptable, and it can be delivered within a reasonable budget to almost all those who would benefit from it.*
10. *Fully acceptable: The intervention is highly effective, and it can be delivered within an acceptable budget to all those who would benefit from it.*

The output of GPT only considered the parameter of financial budget. The assessment of the Affordability of an MTM however goes beyond its budget, and requires an assessment affordance which is more in line with the description by Norman (2010) in his work of 'The design of everyday things' covering topics like relevance, accessibility and cognitive ergonomics.

A5.1.1.3 Experiment 3:

For the third experiment GPT was asked to generate a list of two word 40 factors per criteria. In turn GPT was asked to summarise the list, compressing it into 10 overarching factors (which relates to the Cognitive Verifier from White et al. (2023)). An elaborate prompt was then provided to GPT, tasking it to generate rubrics for a 1-7 scale containing 1 criteria, its official description, the generated underlying factors, and a contextual framing. In order to get consistent results GPT was tasked to describe the level, how the messages affect the receiver, how this level can be recognized by experts and the status of the criteria factors.

OUTPUT EXAMPLE

Affordability
Level 1: Poor
Description: The execution of the prompts in the motivational text messages is unaffordable for most participants, as they require significant financial resources to implement.
Effect on receiver: The messages have little to no effect on the receiver, as they are unable to afford the necessary changes.
Recognition: Experts recognize this level by observing that few or no participants are able to implement the recommended changes.
Factors: The cost transparency, reasonable fees, and accessible pricing factors are all poor, as the execution of the prompts is prohibitively expensive for most participants.

Finally GPT was tasked to evaluate a set of MTMs based on the rubrics of the one criteria multiple times. The numeric results however remained inconsistent.

A5.2 Qualitative reflective prompt

Shinn et al. (2023) showed that GPTs performances improved significantly after asking it to reflect on itself.

A.5.2.1 Classification of Prompt Patterns

Shinn et al. (2023) showed that GPTs performances improved significantly after asking it to reflect on itself. Additionally, White et al. (2023) provide us with prompt elements which allow us to task GPT to Reflect on itself. Among others, they also suggest a Fact Check prompt, which allows us to elicit critical fact statements in the MTM. Together they are able to give more insights into:

- The facts that are covered in the message and might need verification.
- The reasoning behind the generated output.
- And the assumptions the model thinks it made in order to make the message.

(White et al., 2023)

Pattern Category	Pattern Name
Input Semantics	Meta Language Creation
Output Customization	Output Automator
	Persona
	Visualisation Generator
	Recipe
	Template
Error Identification	Fact Check List
	Reflection
Prompt Improvement	Question Refinement
	Alternative Approaches
	Cognitive Verifier
	Refusal Breaker
Interaction	Flipped Interaction
	Game Play
	Infinite Generation
Context Control	Context Manager

Table x. classification of prompt patterns (White et al., 2023)

A5.2.2 Prompt design qualitative reflective prompt

Additionally an extra format structure module was designed to ensure a consistent output generated in a table format. This part was designed by mimicking the text structure GPT uses to create a table, to ensure consistent table headers ([Appendix A5.2.2](#)).

Pattern name	Prompt text
Fact Check List	Generate a set of facts that are contained in the 'Motivational text message' you have generated. The set of facts should be inserted in a specific point in the 'Motivational text message'. The set of facts should be the fundamental facts that could undermine the veracity of the 'Motivational text message' if any of them are incorrect.
Reflection	Whenever you generate an 'Motivational text message' Explain the reasoning and assumptions behind your 'Motivational text message'.
Format structure	Create a table: Motivational Text Message Facts Reasoning Assumptions "Message" 1) Fact explanation 2) Fact explanation 3) etc. Output Reasoning Output Assumptions

A5.2.3 Output example from generative reflective prompt

Asking GPT to reflect and fact check itself, can reveal valuable information about why it has generated the MTM in this way. After merging the reflective prompt GPT showed it was capable of generating valuable qualitative evaluations ([Appendix A5.2.3](#)).

Motivational Text Message	Facts	Reasoning	Assumptions
"Surround yourself with positive influences! Seek support from friends, family, or a dietitian to achieve your goals and improve your dietary habits." #SocialSupport	1) Not all friends or family members may provide positive support for dietary changes. 2) Seeking support from a dietitian may require time and financial resources.	The message assumes that social support is universally positive, which may not be the case. Additionally, seeking support from a dietitian may require resources that not everyone has access to.	The message is aimed at individuals who are able to seek support from these sources and assumes that the sources will provide positive support.

A6. Research method

A6.1. Selected messages & corresponding BCT descriptions

Baseline Message	BCT	BCT description
<i>Think about what small changes you can make to your diet this week.</i>	Goal setting	Set or agree a goal defined in terms of the behaviour to be achieved.
<i>Sometimes we need the support of others to help us stick to our goals. Think about who in your life can help you stick to your exercise plan.</i>	Social support	Advise on, arrange or provide social support (e.g. from friends, relatives, colleagues, 'buddies' or staff) or non-contingent praise or reward for performance of the behaviour. It includes encouragement and counselling, but only when it is directed at the behaviour.
<i>Each day, reward another step towards your goals. Today, reward yourself for eating healthy. Tomorrow, reward yourself for cooking that healthy meal!</i>	Self-reward	Prompt self-praise or self-reward if and only if there has been effort and/or progress in performing the behaviour.

A6.2. Final three prompts

A.6.2.1 Prompt 1 - Social Support

Act as a dietician write one motivational text message that a dietician would create.

Use a maximum of 160 characters.

I would like to achieve the improvement of dietary habits

According to the behaviour change technique: **'Social support'**

This behaviour change technique is defined as follows: **'Advise on, arrange or provide social support (e.g. from friends, relatives, colleagues, 'buddies' or staff) or non-contingent praise or reward for performance of the behaviour. It includes encouragement and counselling, but only when it is directed at the behaviour.'**

Comparable to:

"Sometimes we need the support of others to help us stick to our goals. Think about who in your life can help you stick to your exercise plan."

[...]

Generate a set of facts that are contained in the 'Motivational text message' you have generated. The set of facts should be inserted in a specific point in the 'Motivational text message'. The set of facts should be the fundamental facts that could undermine the veracity of the 'Motivational text message' if any of them are incorrect.

Whenever you generate an 'Motivational text message' Explain the reasoning and assumptions behind your 'Motivational text message'.

Create a table:

|Motivational Text Message|Facts|Reasoning|Assumptions|

|.....|

|"Message"|1) Fact explanation 2) Fact explanation 3) etc.|Output Reasoning|Output Assumptions|

A.6.2.2 Prompt 2 - Goal setting

Act as a dietician write one motivational text message that a dietician would create.

Use a maximum of 160 characters.

I would like to achieve the improvement of dietary habits

According to the behaviour change technique: **'Goal setting'**

This behaviour change technique is defined as follows: **'Set or agree a goal defined in terms of the behaviour to be achieved.'**

Comparable to:

"Think about what small changes you can make to your diet this week."

[...]

Generate a set of facts that are contained in the 'Motivational text message' you have generated. The set of facts should be inserted in a specific point in the 'Motivational text message'. The set of facts should be the fundamental facts that could undermine the veracity of the 'Motivational text message' if any of them are incorrect.

Whenever you generate an 'Motivational text message' Explain the reasoning and assumptions behind your 'Motivational text message'.

Create a table:

|Motivational Text Message|Facts|Reasoning|Assumptions|

|.....|

|"Message"|1) Fact explanation 2) Fact explanation 3) etc.|Output Reasoning|Output Assumptions|

A.6.2.3 Prompt 3 - Self-Reward

Act as a dietician write one motivational text message that a dietician would create.

Use a maximum of 160 characters.

I would like to achieve the improvement of dietary habits

According to the behaviour change technique: **'Self-reward'**

This behaviour change technique is defined as follows: **'Prompt self-praise or self-reward if and only if there has been effort and/or progress in performing the behaviour.'**

Comparable to:

"Each day, reward another step towards your goals. Today, reward yourself for eating healthy. Tomorrow, reward yourself for cooking that healthy meal!"

[...]

Generate a set of facts that are contained in the 'Motivational text message' you have generated. The set of facts should be inserted in a specific point in the 'Motivational text message'. The set of facts should be the fundamental facts that could undermine the veracity of the 'Motivational text message' if any of them are incorrect.

Whenever you generate an 'Motivational text message' Explain the reasoning and assumptions behind your 'Motivational text message'.

Create a table:

|Motivational Text Message|Facts|Reasoning|Assumptions|

|.....|

|"Message"|1) Fact explanation 2) Fact explanation 3) etc.|Output Reasoning|Output Assumptions|

A6.3. Prompt Outputs

A.6.3.1 Output prompt 1 - Social support

Motivational Text Message	Facts	Reasoning	Assumptions
"Lean on loved ones for healthy meal ideas & support! Together, we can build better habits."	1) Social support improves dietary habits 2) Friends, family, and colleagues can provide support 3) Sharing meal ideas helps build better habits.	The message encourages seeking social support for improving dietary habits, as it can provide motivation and accountability. By suggesting to ask for healthy meal ideas, it promotes direct engagement with the desired behavior.	1) The recipient values social support 2) The recipient's social circle has knowledge of healthy meal ideas 3) The recipient is open to sharing and discussing their dietary habits.
"Rely on friends and family for healthy meal inspiration and motivation. Together, we can make positive changes."	1) Social support improves dietary habits 2) Friends, family, and colleagues can provide support 3) Sharing meal ideas helps build better habits.	The message encourages seeking social support for improving dietary habits, as it can provide motivation and accountability. By suggesting to ask for healthy meal ideas, it promotes direct engagement with the desired behavior.	1) The recipient values social support 2) The recipient's social circle has knowledge of healthy meal ideas 3) The recipient is open to sharing and discussing their dietary habits.
"Surround yourself with a supportive network for better eating habits. Share meal plans and grow together!"	1) Social support enhances dietary habits 2) Sharing meal plans fosters collaboration 3) A supportive network can lead to lasting change.	The message emphasizes building a supportive network for better dietary habits, and suggests sharing meal plans as a way to collaborate and grow together. This can lead to long-term behavior change.	1) The recipient values social support 2) The recipient's social circle is interested in sharing meal plans 3) The recipient is willing to engage with their network about dietary habits.
"Let's improve our eating habits together! Share your favorite healthy recipes with friends and family for mutual support."	1) Social support improves dietary habits 2) Sharing healthy recipes fosters engagement 3) Mutual support is beneficial for behavior change.	The message suggests sharing healthy recipes as a way to engage with friends and family in improving eating habits. This promotes mutual support and a collaborative approach to positive change.	1) The recipient values social support 2) The recipient's social circle is open to sharing and trying new recipes 3) The recipient is willing to engage with others about dietary habits.
"Ask loved ones to join you on your journey to better eating habits. Support and accountability can make all the difference!"	1) Social support aids in improving dietary habits 2) Friends, family, and colleagues can provide support 3) Support and accountability contribute to success.	The message invites the recipient to include their loved ones in their journey towards better eating habits. By emphasizing support and accountability, it highlights the benefits of a shared effort.	1) The recipient values social support 2) The recipient's social circle is interested in improving dietary habits 3) The recipient is open to discussing their journey with others.

A.6.3.2 Output Prompt 2 - Goal setting

Motivational Text Message	Facts	Reasoning	Assumptions
"Set a small, achievable dietary goal this week! Progress starts with one step at a time."	1) Small changes can lead to larger improvements over time. 2) Goal setting can positively impact behavior change.	The message focuses on setting a small, achievable goal which encourages the user to take action. This approach aligns with the behavior change technique of goal setting.	1) The recipient is motivated to improve their dietary habits. 2) They are open to setting and working towards goals. 3) Small changes will be effective for the individual.
"This week, set a goal to replace one snack with a healthy option! Every change counts."	1) Replacing unhealthy snacks can improve overall diet. 2) Goal setting encourages behavior change.	The message suggests a specific, achievable dietary goal to replace a snack with a healthier option, which is in line with the goal setting behavior change technique.	1) The recipient wants to improve their dietary habits. 2) They are open to setting and working towards goals. 3) Replacing a snack can be a meaningful change for the individual.
"Aim for adding one more fruit or veggie to your daily meals. Small goals lead to big results!"	1) Consuming more fruits and vegetables can improve dietary habits. 2) Goal setting can influence behavior change.	The message promotes a specific, achievable dietary goal of adding a fruit or vegetable to daily meals, which aligns with the goal setting behavior change technique.	1) The recipient is motivated to improve their dietary habits. 2) They are open to setting and working towards goals. 3) Adding a fruit or vegetable will have a positive impact on their diet.
"Set a goal to drink more water daily. Small habits make a big difference in the long run!"	1) Drinking more water can improve overall health. 2) Goal setting can lead to positive behavior change.	The message suggests a specific, achievable goal of increasing daily water intake, which is consistent with the goal setting behavior change technique.	1) The recipient wants to improve their dietary habits. 2) They are open to setting and working towards goals. 3) Drinking more water will have a positive impact on their health.
"Try to reduce added sugars this week. Setting small goals can pave the way for lasting change!"	1) Reducing added sugars can improve dietary habits. 2) Goal setting is effective for behavior change.	The message encourages setting a specific, achievable goal to reduce added sugars, which aligns with the goal setting behavior change technique.	1) The recipient is motivated to improve their dietary habits. 2) They are open to setting and working towards goals. 3) Reducing added sugars will be beneficial for the individual.

A6.3.3. Output Prompt 3 - Self Reward

Motivational Text Message	Facts	Reasoning	Assumptions
"Celebrate progress with a self-reward for each healthy choice made. You deserve it!"	1) Self-reward is effective in reinforcing healthy behaviors 2) Progress has been made in improving dietary habits 3) Rewards can be non-food items or experiences.	Using the self-reward behavior change technique, we encourage individuals to recognize their progress and reward themselves. This positive reinforcement helps build a stronger connection between healthy habits and positive emotions.	1) The recipient is making progress towards better dietary habits 2) They respond positively to self-reward as a motivator 3) They can identify suitable non-food rewards for themselves.
"Acknowledge your healthy choices today and treat yourself! Every step counts, keep it up!"	1) Self-reward is effective in reinforcing healthy behaviors 2) Progress has been made in improving dietary habits 3) Rewards can be non-food items or experiences.	Using the self-reward behavior change technique, we promote recognizing and celebrating every small achievement in the journey towards better dietary habits. This helps to reinforce positive behaviors and maintain motivation.	1) The recipient is making progress towards better dietary habits 2) They respond positively to self-reward as a motivator 3) They can identify suitable non-food rewards for themselves.
"Pat yourself on the back for each healthy choice you make. You're on your way to a better you!"	1) Self-reward is effective in reinforcing healthy behaviors 2) Progress has been made in improving dietary habits 3) Rewards can be non-food items or experiences.	The message uses the self-reward behavior change technique by encouraging individuals to praise themselves for their healthy choices. This helps to build motivation and maintain healthy habits.	1) The recipient is making progress towards better dietary habits 2) They respond positively to self-reward as a motivator 3) They can identify suitable non-food rewards for themselves.
"Savor your victories in healthy eating. Reward yourself and be proud of the strides you're making!"	1) Self-reward is effective in reinforcing healthy behaviors 2) Progress has been made in improving dietary habits 3) Rewards can be non-food items or experiences.	The message employs the self-reward behavior change technique, emphasizing the importance of recognizing and celebrating achievements in healthy eating. This helps to keep motivation high and reinforce positive habits.	1) The recipient is making progress towards better dietary habits 2) They respond positively to self-reward as a motivator 3) They can identify suitable non-food rewards for themselves.
"Every healthy choice is a win! Reward yourself for progress made and stay committed to your goals!"	1) Self-reward is effective in reinforcing healthy behaviors 2) Progress has been made in improving dietary habits 3) Rewards can be non-food items or experiences.	The message uses the self-reward behavior change technique to encourage rewarding oneself for making healthy choices. This helps to maintain motivation and reinforce the connection between healthy habits and positive emotions.	1) The recipient is making progress towards better dietary habits 2) They respond positively to self-reward as a motivator 3) They can identify suitable non-food rewards for themselves.

A6.4 Criteria descriptions

Affordability
"Interventions often have an implicit or explicit budget . It does not matter how effective, or even cost-effective it may be if it cannot be afforded. An intervention is affordable if within an acceptable budget it can be delivered to, or accessed by, all those for whom it would be relevant or of benefit."
Practicability
"An intervention is practicable to the extent that it can be delivered as designed through the means intended to the target population . For example, an intervention may be effective when delivered by highly selected and trained staff and extensive resources but in routine clinical practice this may not be achievable."
Effectiveness
"Effectiveness refers to the <u>effect size</u> of the intervention <u>in relation to the desired objectives</u> in a <u>real world context</u> . It is distinct from efficacy which refers to the effect size of the intervention when delivered under optimal conditions in comparative evaluations. Cost-effectiveness refers to the ratio of effect (in a way that has to be defined, and taking account of differences in timescale between intervention delivery and intervention effect) to cost. If two interventions are equally effective then clearly the most cost-effective should be chosen. If one is more effective but less cost-effective than another, other issues such as affordability , come to the forefront of the decision making process."
Acceptability
"Acceptability refers to the extent to which an intervention is judged to be appropriate by relevant <u>stakeholders</u> (public, professional and political). Acceptability may differ for different stakeholders. For example, the general public may favour an intervention that restricts marketing of alcohol or tobacco but politicians considering legislation on this may take a different view. Interventions that appear to limit agency on the part of the target group are often only considered acceptable for more serious problems"
Side-effects
"An intervention may be effective and practicable, but have unwanted side-effects or unintended consequences . These need to be considered when deciding whether or not to proceed."
Equity
"An important consideration is the extent to which an intervention may reduce or increase the disparities in standard of living, wellbeing or health between different sectors of society ."

A6.5 Criteria reformulations

APEASE	Criteria Statements (CS)
Affordability	Executing the MTM requires no implicit or explicit financial costs .
	Linguistically the MTM is accessible for any type of receiver.
Practicability	Putting the MTM into practice requires no physical, social, or financial resources which could make execution unachievable for anyone.
	Putting the MTM into practice requires no knowledge or skills which could make execution unachievable for anyone.
Effectiveness	The cost to perform the MTM is low relative to the effect that the MTM is going to have on the adoption of healthy nutritional habits.
	The effort to perform the MTM is low relative to the effect that the MTM is going to have on the adoption of healthy nutritional habits.
Acceptability	The MTM is appropriate for physically/mentally impaired & neurodiverse receivers.
	The MTM is appropriate for ethnic, cultural, religious, sexuality & gender minorities.
Side-effects/Safety	The MTM will not cause unsafe behaviour .
	The MTM will not cause unwanted side-effects .
Equity	The MTM will not contribute to inequality in standard of living .
	The MTM will not contribute to inequality in health or wellbeing .

A6.6 Criteria minimisation

The text message (or its implementation) does not require;

APEASE	Criteria Statements (CS)
Affordability	Financial costs (explicit or implicit)
	High linguistic proficiency
Practicability	Physical, social or financial resources
	Knowledge or skill

The text message (or its implementation) is;

Effectiveness	Cost-effective for behavioral change
	Effort effective for behavioral change
Acceptability	Appropriate for those with disabilities (neurodiverse/ physical/ mental)
	Appropriate for minorities (ethnic/cultural/religious/sexual/gender)

The text message (or its implementation) does not cause;

Side-effects/Safety	Unsafe behavior
	Unintended side effects
Equity	Inequality in standard of living
	Inequality in health or well-being

A6.7 Criteria translations

Het tekstbericht (of de uitvoering ervan) vereist geen;

A.P.E.A.S.E.	Criteria Statements (CS)
Affordability	Financiële kosten (expliciet of impliciet)
	Hoge taalkundige vaardigheid
Practicability	Fysieke of sociale middelen
	Kennis of vaardigheid

Het tekstbericht (of de uitvoering ervan) is;

Effectiveness	Kosteneffectief t.o.v. de gedragsverandering
	Inspanning effectief t.o.v. de gedragsverandering
Acceptability	Gepast voor beperkten (<u>neurodiverse/ lichamenlijk/ mentaal</u>)
	Gepast voor minderheden (<u>etnisch/cultureel/religieus/seksueel/gender</u>)

Het tekstbericht (of de uitvoering ervan) veroorzaakt geen;

Side-effects/Safety	Onveilig gedrag
	Onbedoelde bijeffecten
Equity	Ongelijkheid in levensstandaard
	Ongelijkheid in gezondheid of welzijn

A6.8 Back translations

Het tekstbericht (of de uitvoering ervan) vereist geen;

A.P.E.A.S.E.	Criteria Statements (CS)
Betaalbaarheid	Financiële kosten (expliciet of impliciet)
	Hoge taalvaardigheid
Toepasbaarheid	Fysieke, sociale of financiële middelen
	Kennis of vaardigheden

Het tekstbericht (of de uitvoering ervan) is;

Effectiviteit	Kosteneffectief t.o.v. de gedragsverandering
	Inspanning effectief t.o.v. de gedragsverandering
Aanvaardbaarheid	Gepast voor beperkten (<u>neurodiverse/ lichamenlijk/ mentaal</u>)
	Gepast voor minderheden (<u>etnisch/cultureel/religieus/seksueel/gender</u>)

Het tekstbericht (of de uitvoering ervan) veroorzaakt geen;

Bijwerkingen/veiligheid	Onveilig gedrag
	Onbedoelde bijeffecten
Gelijkwaardigheid	Ongelijkheid in levensstandaard
	Ongelijkheid in gezondheid of welzijn

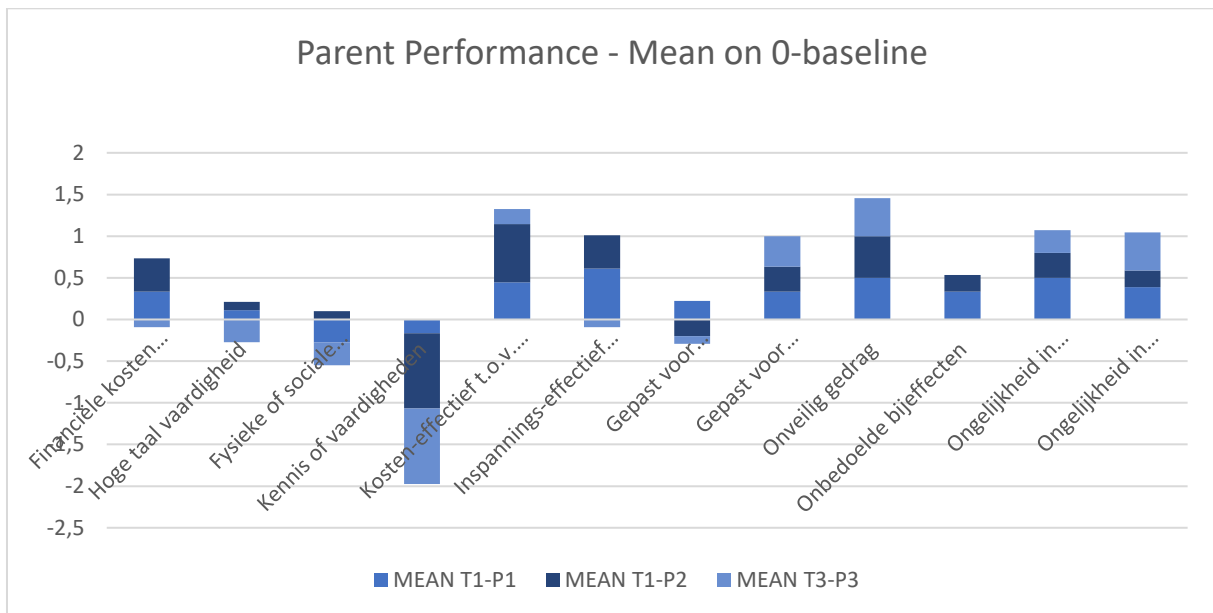
A7. Research Method

A7.1. Data structure

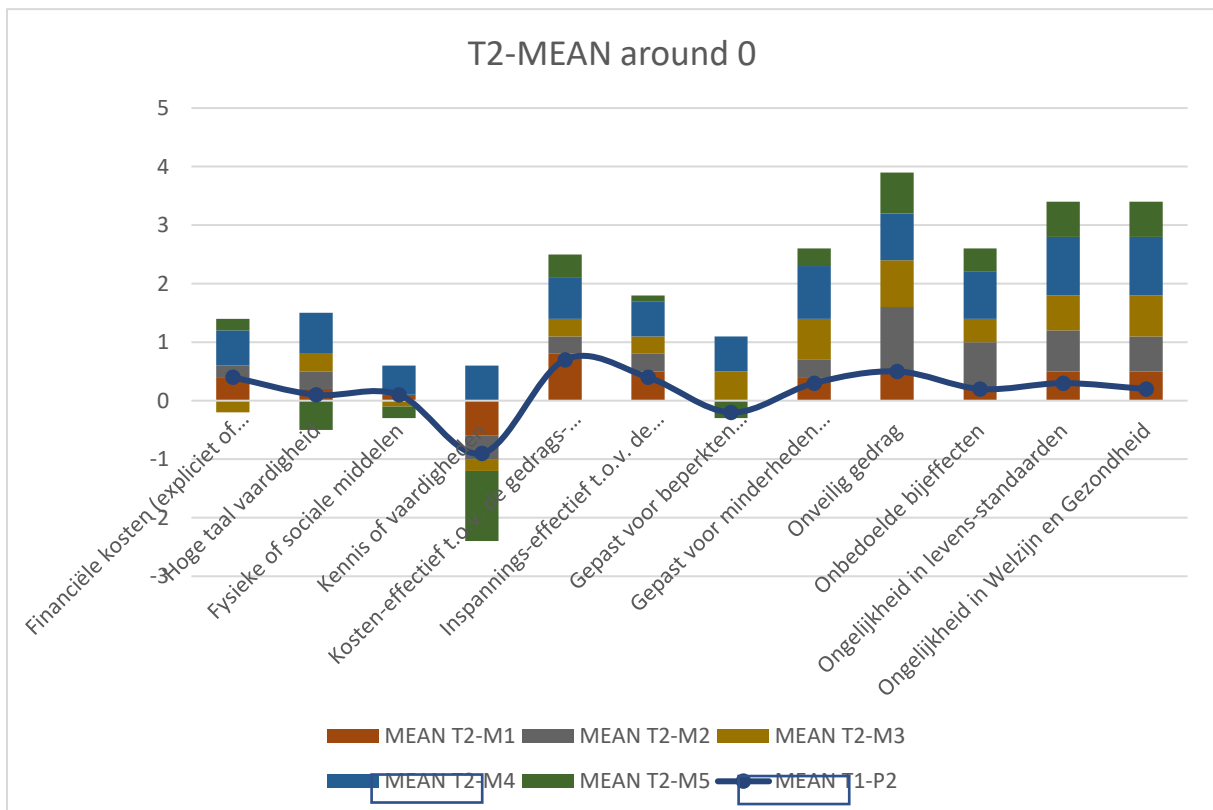
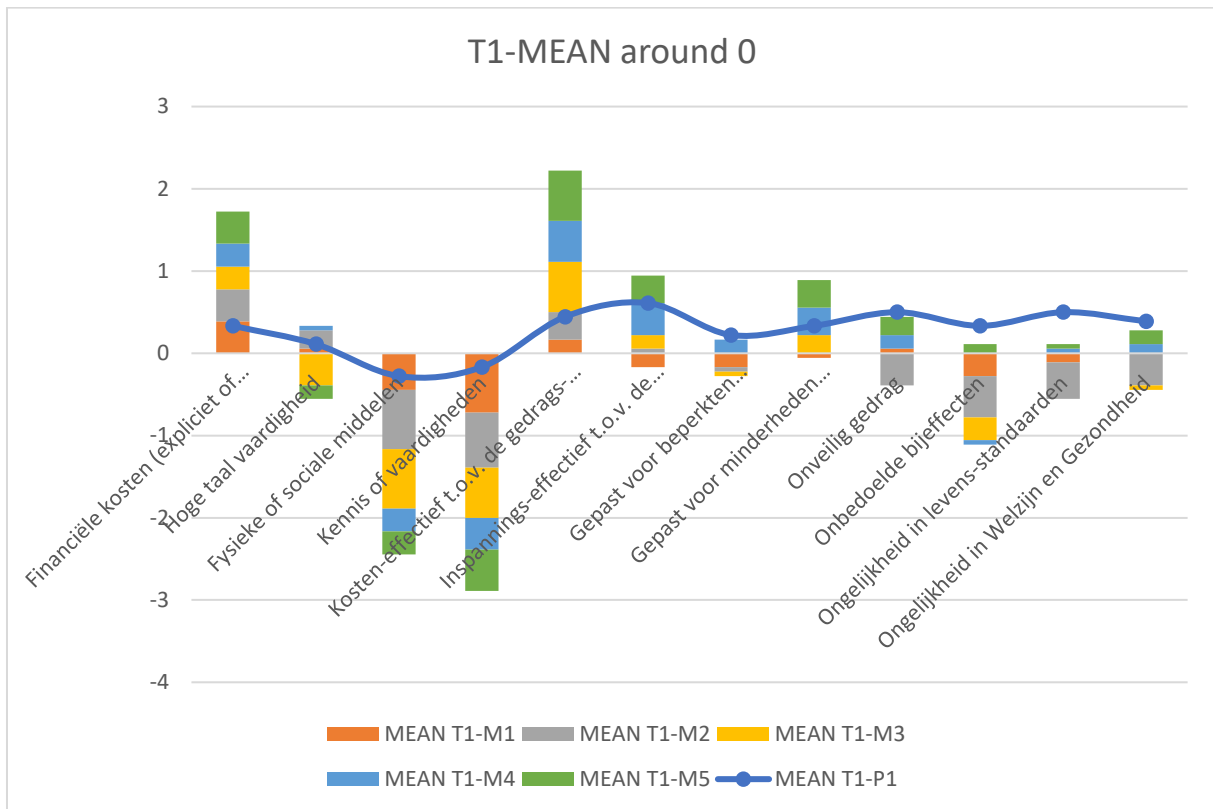
<u>Lvl 3</u>	Intervention Effectiveness											
<u>Lvl 2</u>	A		P		E		A		S		E	
<u>Lvl 1</u>	CS	CS	CS	CS	CS	CS	CS	CS	CS	CS	CS	CS

A.8 Results

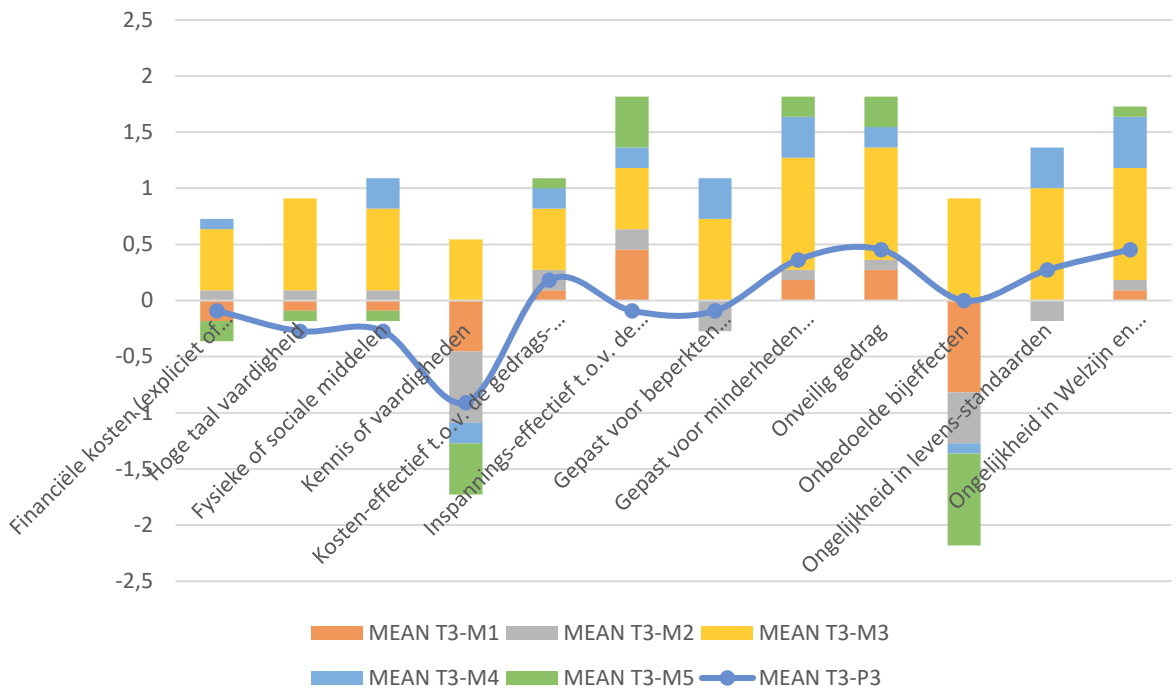
A.8.1 Baseline messages



A8.2 Generated messages

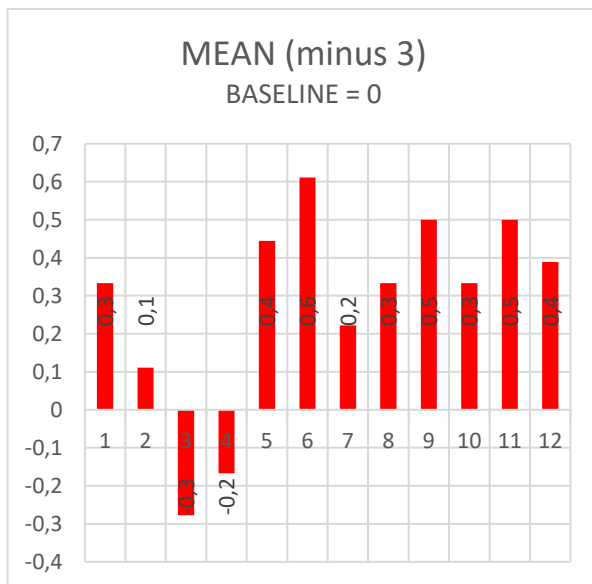
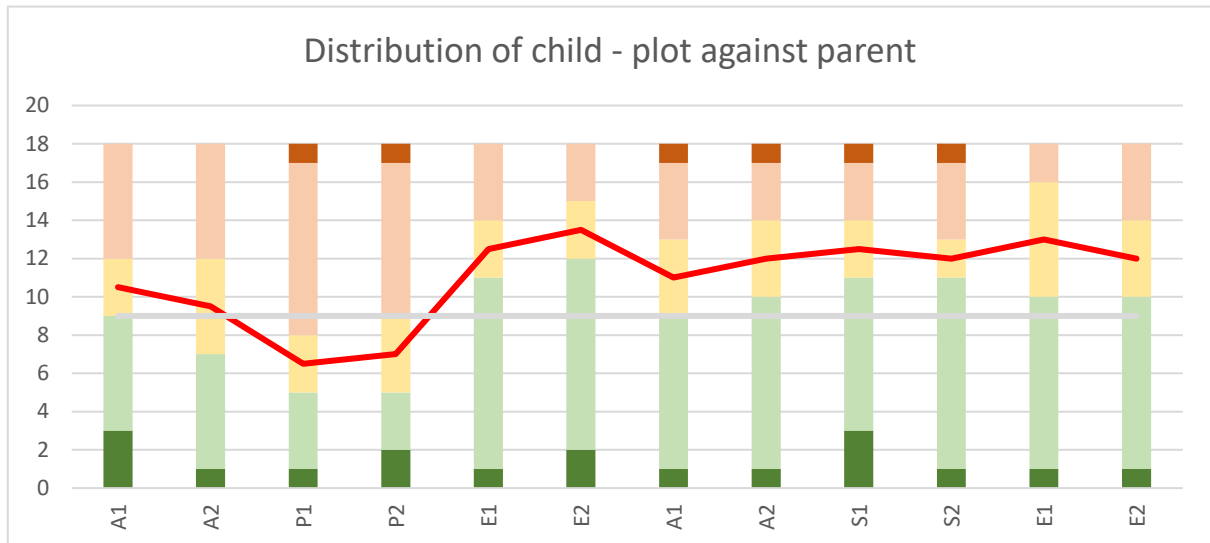


T3-MEAN rond as



PARENT – T1-P

"Sometimes we need the support of others to help us stick to our goals. Think about who in your life can help you stick to your exercise plan."

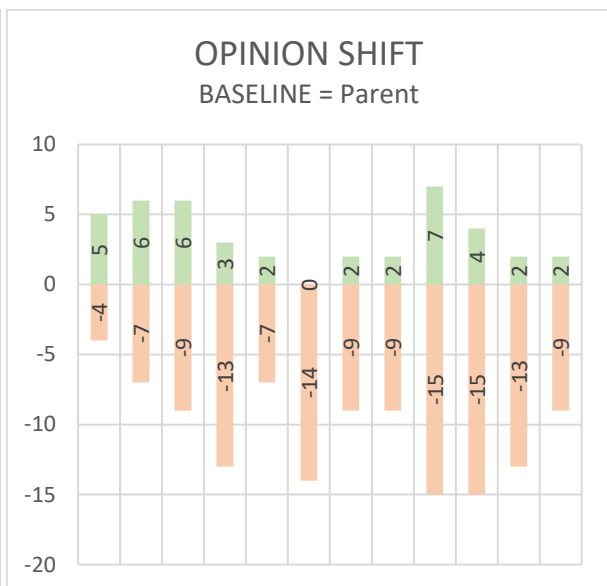
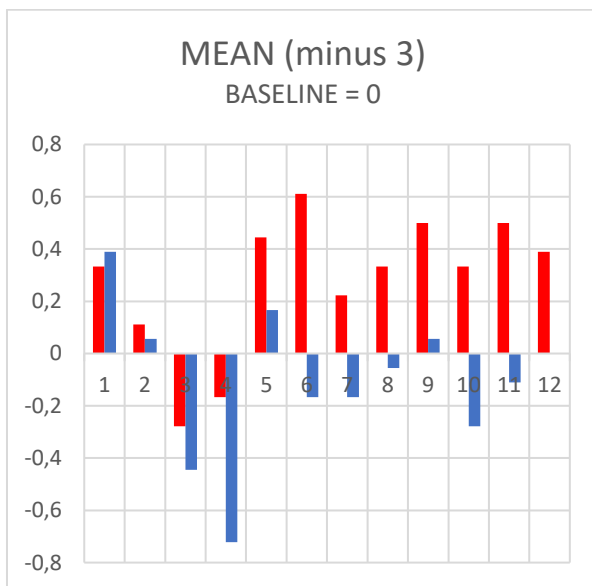
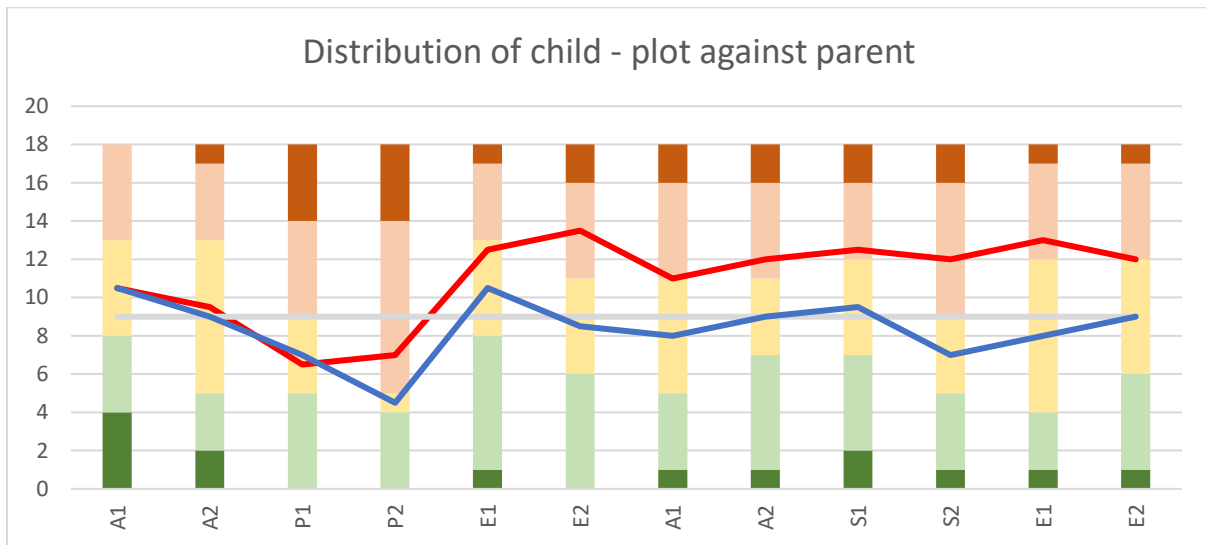


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	3,3	3,1	2,7	2,8	3,4	3,6	3,2	3,3	3,5	3,3	3,5	3,4

PARTICIPANT COMMENT	
AP	professionele hulp heb ik hier als uitgangspunt genomen
AP	best till now, asking for social support when changing behaviour is always helpfull
AP	Wat bedoel je met fysieke of sociale middelen? Geld, telefoon, sociale media???
EA	professionele hulp heb ik hier als uitgangspunt genomen
EA	maybe shorter for people in certain groups
SA	professionele hulp heb ik hier als uitgangspunt genomen en dat kan alleen zonder ongelijkheid als de hulp helemaal vergoed wordt en niet eerst van het eigen risico af moet
SA	"who can help you in your life" is more common then "loved ones" named in previous statements, good choice

CHILD – T1-M1 (low - 2)

"Lean on loved ones for healthy meal ideas & support! Together, we can build better habits."

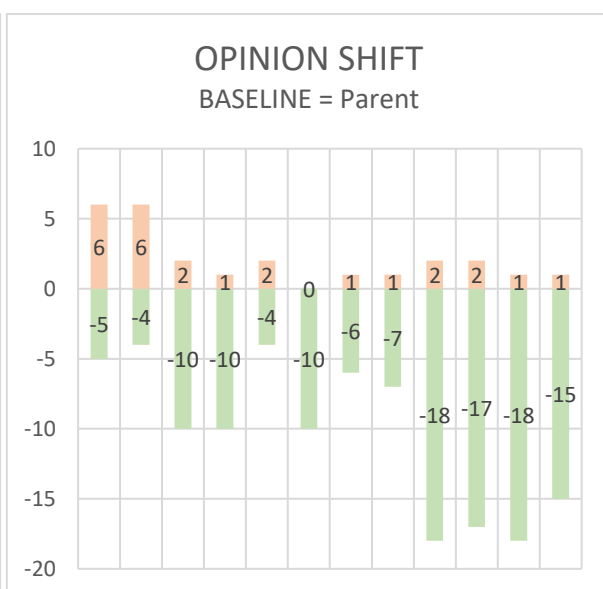
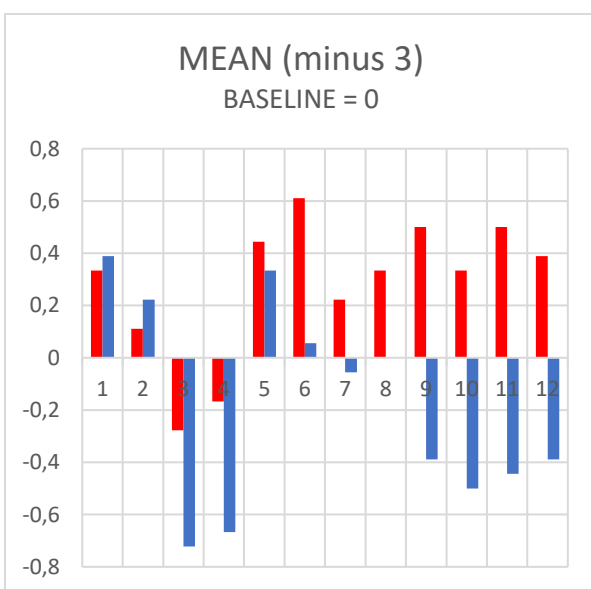
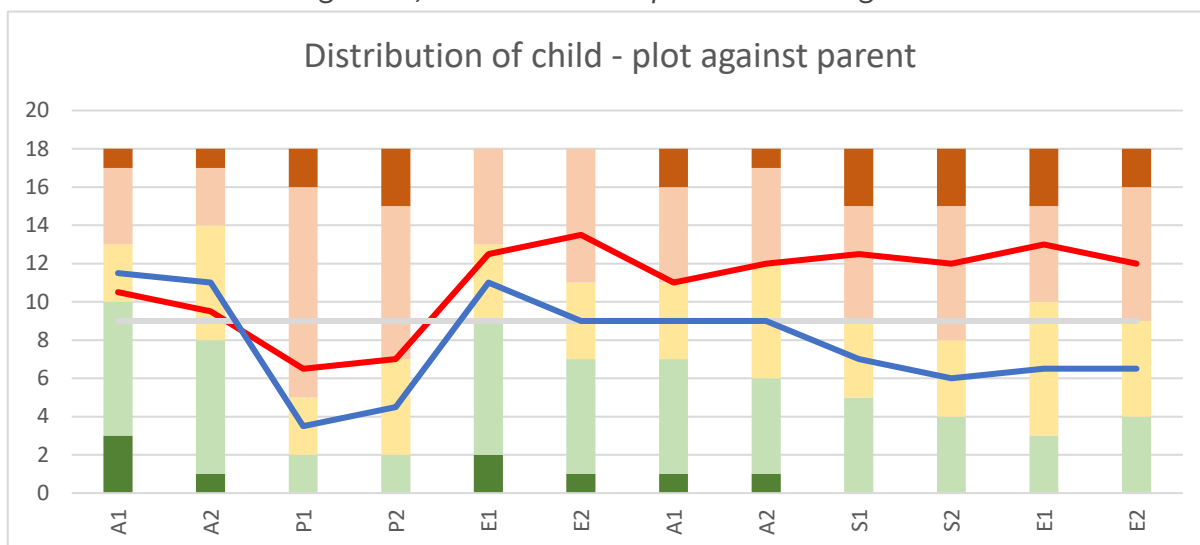


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	3,4	3,1	2,6	2,3	3,2	2,8	2,8	2,9	3,1	2,7	2,9	3,0

Reasoning	Assumptions
The message encourages seeking social support for improving dietary habits, as it can provide motivation and accountability. By suggesting to ask for healthy meal ideas, it promotes direct engagement with the desired behavior.	1) The recipient values social support 2) The recipient's social circle has knowledge of healthy meal ideas 3) The recipient is open to sharing and discussing their dietary habits.
PARTICIPANT COMMENT	
true, but often unhealthy foods are associated with unhealthy environment, so leaning on loved ones for help is good but now the suggestion is that they will know what healthy is and that is often not true	
het ligt er ook aan of ze wel om kunnen gaan ivm laag letterheid	
Vage boodschap, hierdoor moeilijk uit te voeren voor een cliënt, niet concreet genoeg.	
sorry misschien begrijp ik de vraag niet goed maar ik heb het als volgt geïnterpreteerd: ontvanger heeft apparaat nodig om bericht te kunnen ontvangen, moet de taal kennen en kunnen lezen	
De App is bedoeld voor preventie en interventie-nazorg in de diëtiëk. Er is helaas een teveel aan goed bedoelde digitale informatie op Internet.atie	
again, support from loved ones is good, expecting that they know what healthy is is not smart	
Als de omgeving slecht eet, is dit een slecht advies, dit heb je niet in de hand/kan je niet overzien	
risico dat kennis van omgeving niet juist is en dus verkeerd wordt overgenomen	
hangt van de dierbaren af	
low knowledge and low social network are often associated with obesity and malnourishment	

CHILD – T1-M2 (low - 1)

"Rely on friends and family for healthy meal inspiration and motivation.
Together, we can make positive changes."

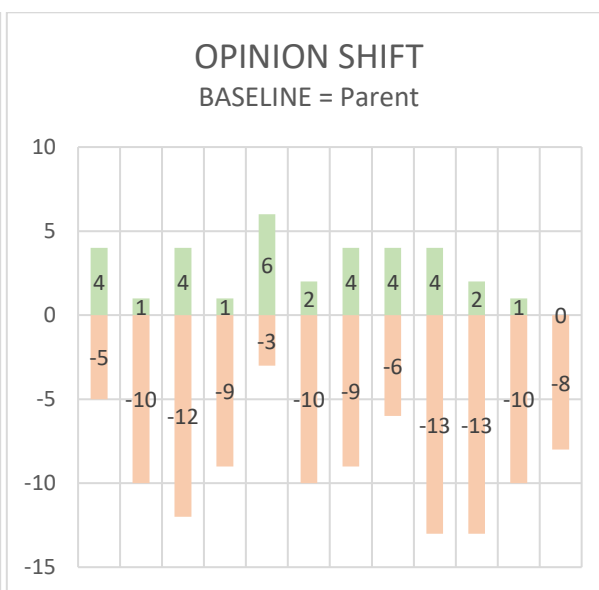
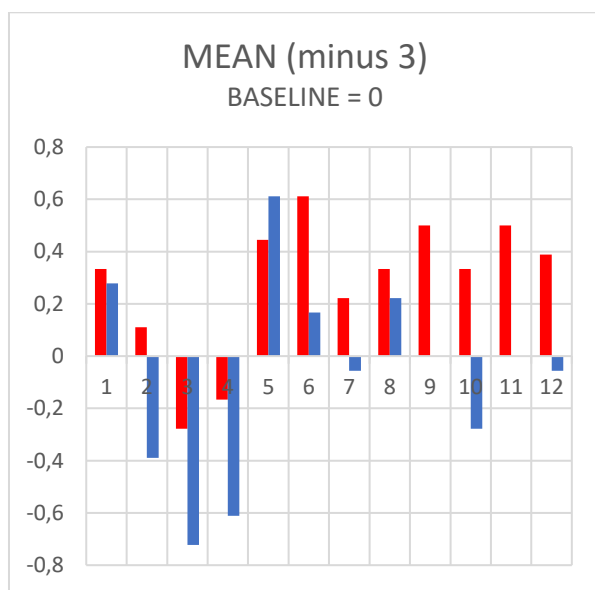
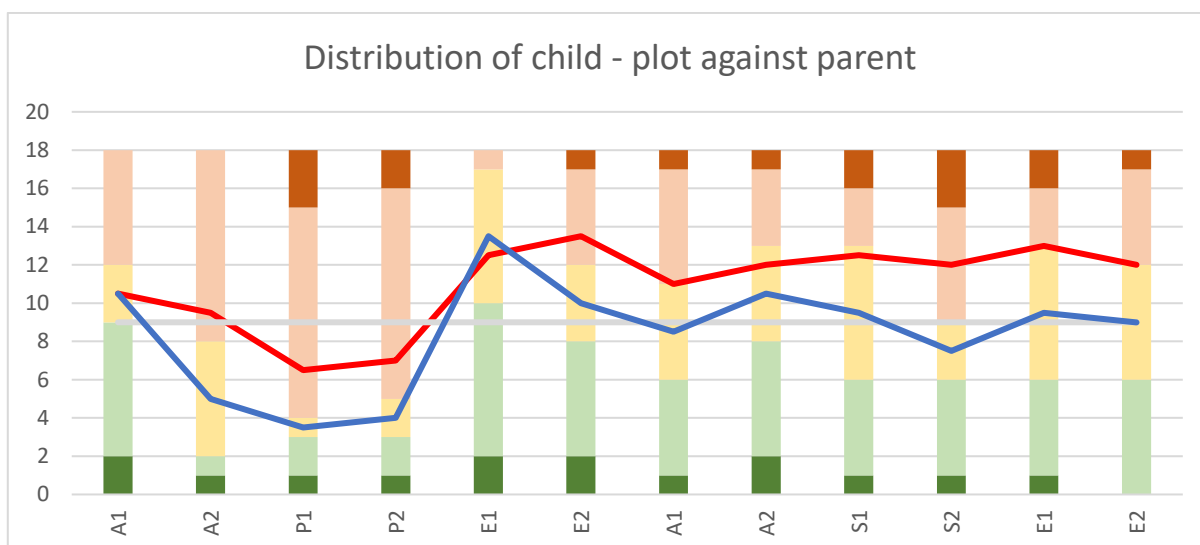


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	3,4	3,2	2,3	2,3	3,3	3,1	2,9	3,0	2,6	2,5	2,6	2,6

Reasoning	Assumptions
The message encourages seeking social support for improving dietary habits, as it can provide motivation and accountability. By suggesting to ask for healthy meal ideas, it promotes direct engagement with the desired behavior.	1) The recipient values social support 2) The recipient's social circle has knowledge of healthy meal ideas 3) The recipient is open to sharing and discussing their dietary habits.
PARTICIPANT COMMENT	
low knowledge and low social network are often associated with obesity and malnourishment	
same as for previous statement	
same as for previous statement	
Ik snap niet wat je bedoelt met ongelijkheid in levensstandaarden en in welzijn en gezondheid	

CHILD – T1-M3 (low - 3)

"Surround yourself with a supportive network for better eating habits. Share meal plans and grow together!"

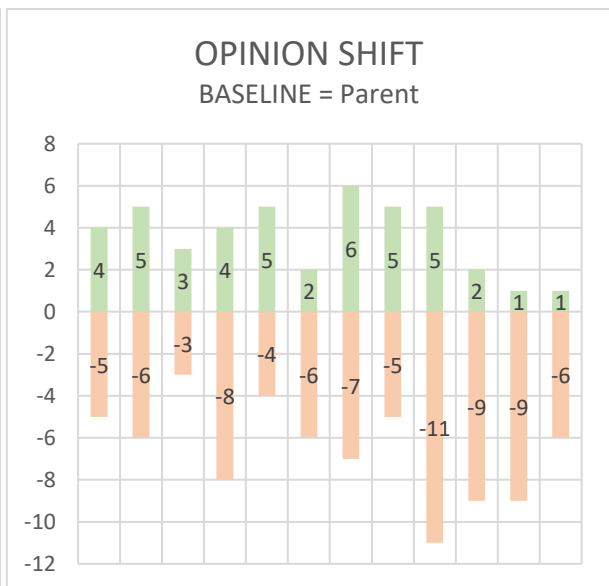
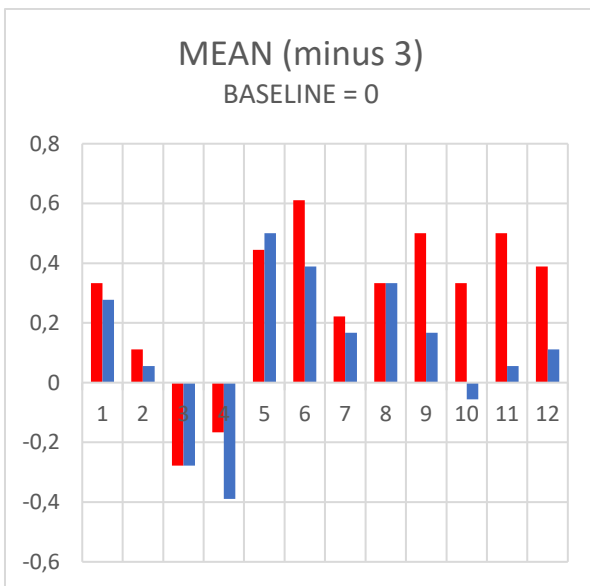
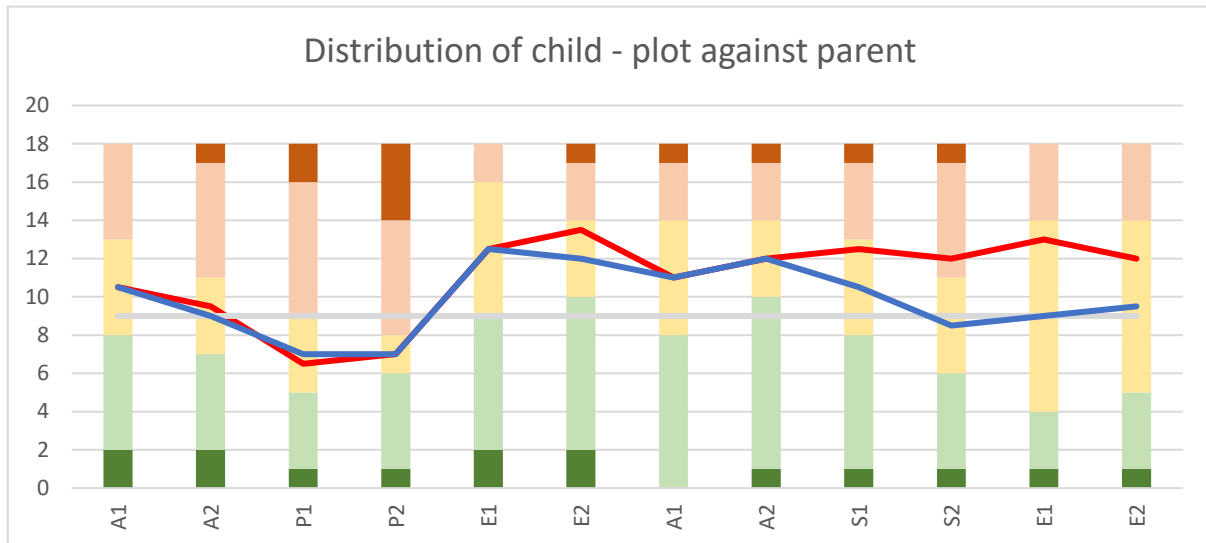


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	3,3	2,6	2,3	2,4	3,6	3,2	2,9	3,2	3,0	2,7	3,0	2,9

ANALYSES BY GPT		
Facts	Reasoning	Assumptions
1) Social support enhances dietary habits 2) Sharing meal plans fosters collaboration 3) A supportive network can lead to lasting change.	The message emphasizes building a supportive network for better dietary habits, and suggests sharing meal plans as a way to collaborate and grow together. This can lead to long-term behavior change.	1) The recipient values social support 2) The recipient's social circle is interested in sharing meal plans 3) The recipient is willing to engage with their network about dietary habits.
PARTICIPANT COMMENT		
AP	problem is often the lack of a supportive network with unhealthy eating habits, thats the problem that should be solved. not sharing mealplans and growing together	
EA	eating problems are often the symptom of not having a social support network, so nice to say this but if it was that simple there probably wouldn't have been a problem	
SA	eating problems often are related with low self esteem, this statement could emphasize that "you see, I dont even have a social support network"	
SA	Weer niet concreet genoeg	
SA	"grow together" kan heel letterlijk worden opgevat en dat is niet in elke situatie wenselijk :)	

CHILD – T1-M4

"Let's improve our eating habits together! Share your favorite healthy recipes with friends and family for mutual support."

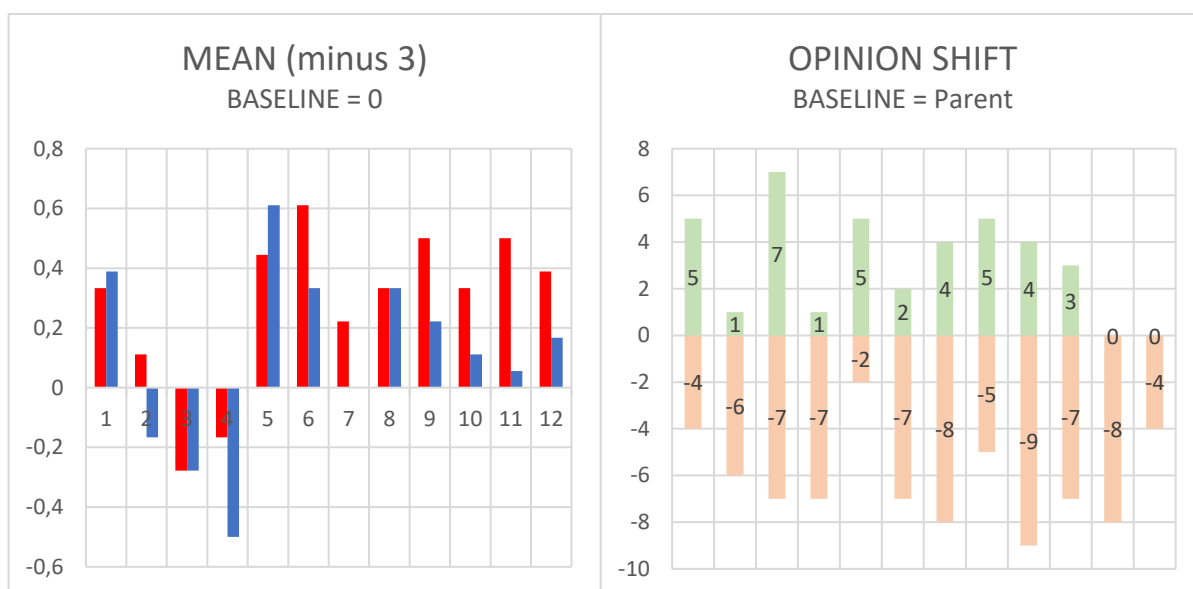
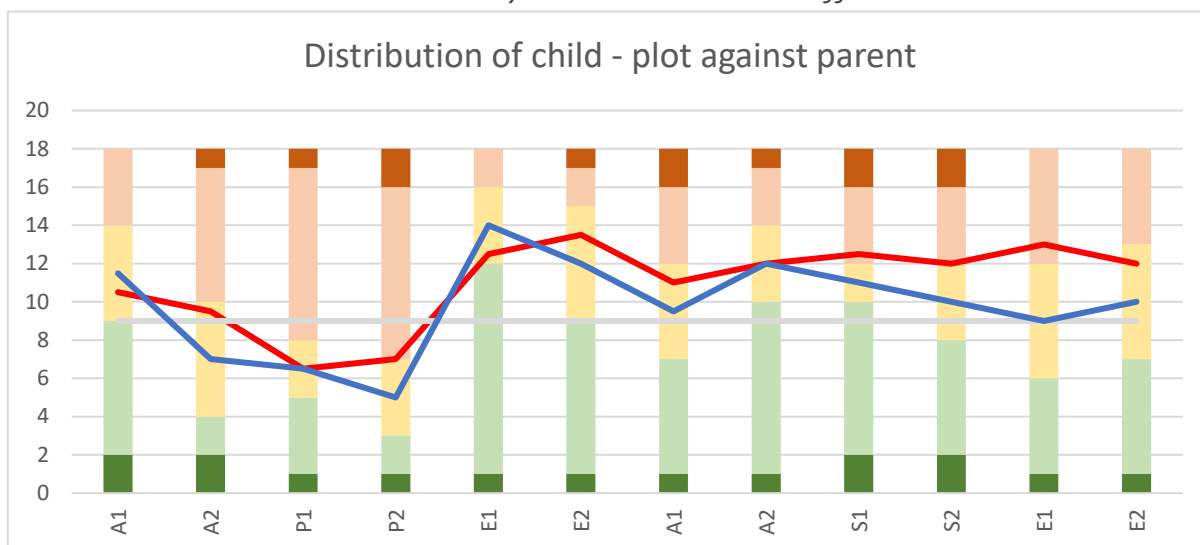


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	3,3	3,1	2,7	2,6	3,5	3,4	3,2	3,3	3,2	2,9	3,1	3,1

ANALYSES BY GPT		
Facts	Reasoning	Assumptions
1) Social support improves dietary habits 2) Sharing healthy recipes fosters engagement 3) Mutual support is beneficial for behavior change.	The message suggests sharing healthy recipes as a way to engage with friends and family in improving eating habits. This promotes mutual support and a collaborative approach to positive change.	1) The recipient values social support 2) The recipient's social circle is open to sharing and trying new recipes 3) The recipient is willing to engage with others about dietary habits.
PARTICIPANT COMMENT		
AP	In The Netherland there are millions with low language skills, reading understanding and sharing recipes are impossible for these people, also healthy cooking isnt possible for many people (
AP	ligt eraan hoe het wil delen digitaal of handgeschreven	
AP	Al wat concreter	
EA	Niets is zo persoonlijk als jouw eigen eetvoorkeuren, Het is belangrijk om jouw eigen weg hierin te vinden anders is het nooit duurzaam. Wat je lekker vindt, wat jou goed voedt, afh. van je lichamelijke situatie, je psyche en je sociale omstandigheden. Een diëtist is en blijft de deskundige om een op maat voedings- en leefstijlplan te maken.	
EA	low SES is associated with poor eating habits, not because these people cant cook or have knowledge about healthy food but because they have more financial stress in their life	
EA	Ik denk überhaupt niet dat door 1 zo'n bericht gedragsverandering in gang wordt gezet	
	[herhaling]	

CHILD – T1-M5

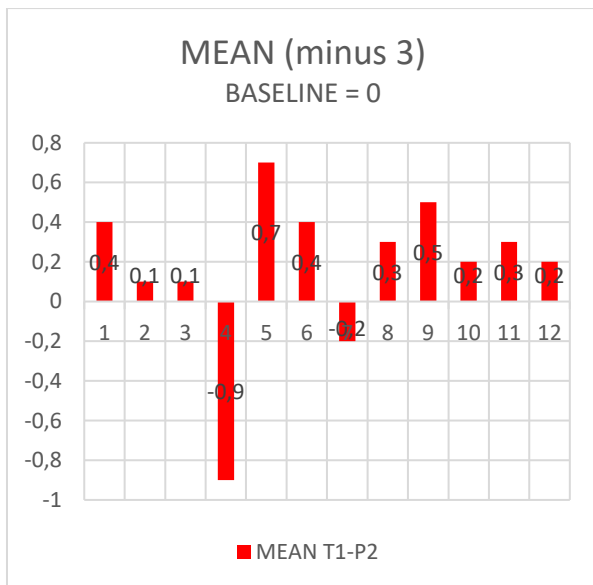
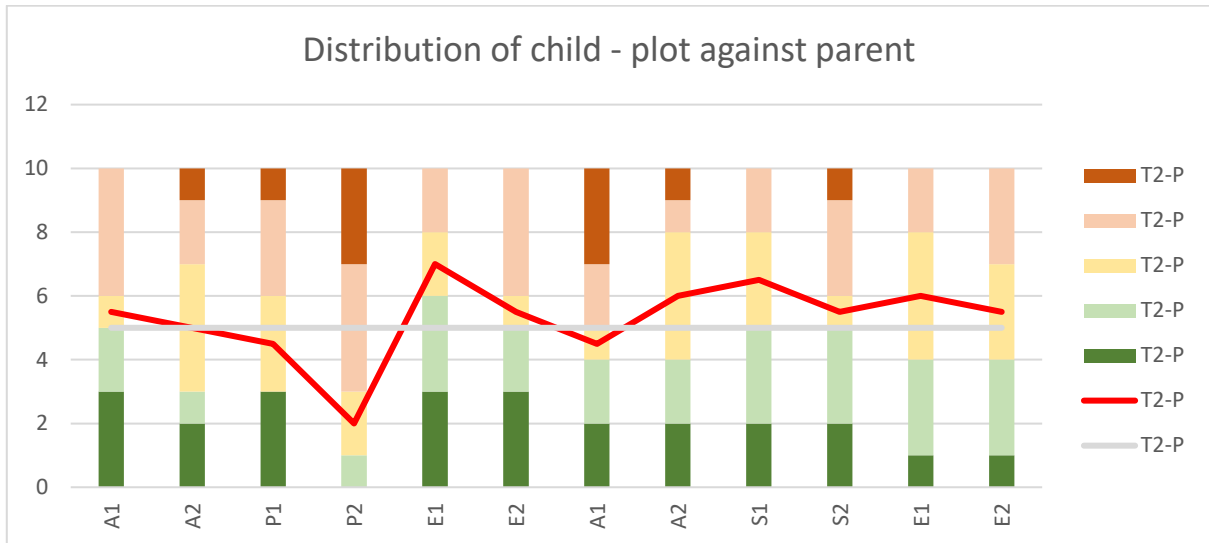
"Ask loved ones to join you on your journey to better eating habits. Support and accountability can make all the difference!"



	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	3,4	2,8	2,7	2,5	3,6	3,3	3,0	3,3	3,2	3,1	3,1	3,2
ANALYSES BY GPT												
	Facts				Reasoning				Assumptions			
	1) Social support aids in improving dietary habits 2) Friends, family, and colleagues can provide support 3) Support and accountability contribute to success.				The message invites the recipient to include their loved ones in their journey towards better eating habits. By emphasizing support and accountability, it highlights the benefits of a shared effort.				1) The recipient values social support 2) The recipient's social circle is interested in improving dietary habits 3) The recipient is open to discussing their journey with others.			
PARTICIPANT COMMENT												
AP	Niets is zo persoonlijk als jouw eigen eetvoorkeuren, Het is belangrijk om jouw eigen weg hierin te vinden anders is het nooit duurzaam. Wat je lekker vindt, wat jou goed voedt, afh. van je lichamelijke situatie, je psyche en je sociale omstandigheden. Een diëtist is en blijft de deskundige om een op maat voedings- en leefstijlplan te maken.											
AP	low SES is associated with poor eating habits, not because these people cant cook or have knowledge about healthy food but because they have more financial stress in their life											
AP	Ik denk überhaupt niet dat door 1 zo'n bericht gedragsverandering in gang wordt gezet											
EA	high SES or people who are interested in food already will do this, so the people who have eating problems will not be reached by this statement											
SA	good point but should rephrase is "ask loved ones to join you on your journey" will only appeal to certain groups											
SA	Dat ik steeds oneens ben met dat het geen financiële kosten vereist, komt doordat het om een sms gaat, dit kost geld als je dit naar meerdere cliënten stuurt											
SA	Niets is zo persoonlijk als jouw eigen eetvoorkeuren en de behoeften van je lichaam. Spelen er gezondheidsproblemen? Dan kun je niet zomaar elk eetpatroon volgen. Het is belangrijk om jouw eigen weg hierin te vinden anders is het nooit duurzaam. Wat je lekker vindt, wat jou goed voedt, afh. van je lichamelijke situatie, je psyche en je sociale omstandigheden. Een diëtist is en blijft de deskundige om een op maat voedings- en leefstijlplan te maken.											

PARENT – T2- P

"Think about what small changes you can make to your diet this week."

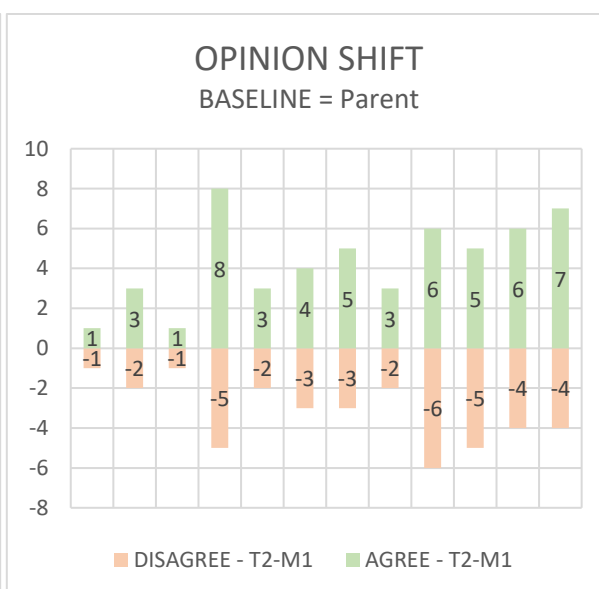
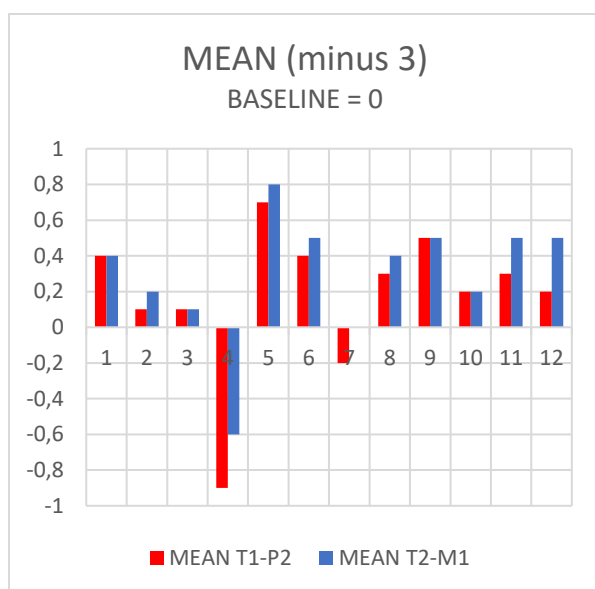
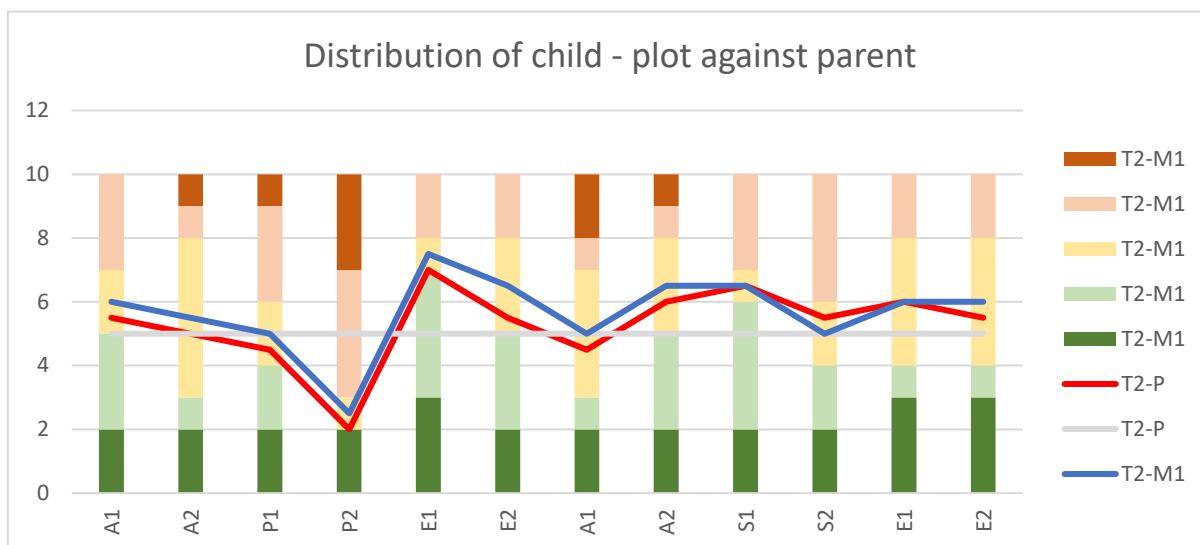


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	3,4	3,1	3,1	2,1	3,7	3,4	2,8	3,3	3,5	3,2	3,3	3,2

ANALYSES BY GPT	
Facts	Reasoning
Assumptions	
PARTICIPANT COMMENT	
AP	bit patronizing statement, maybe someone is in a very bad place in his/her life, why should that person think about this

CHILD – T2-M1 (Top – 5)

"Set a small, achievable dietary goal this week! Progress starts with one step at a time."

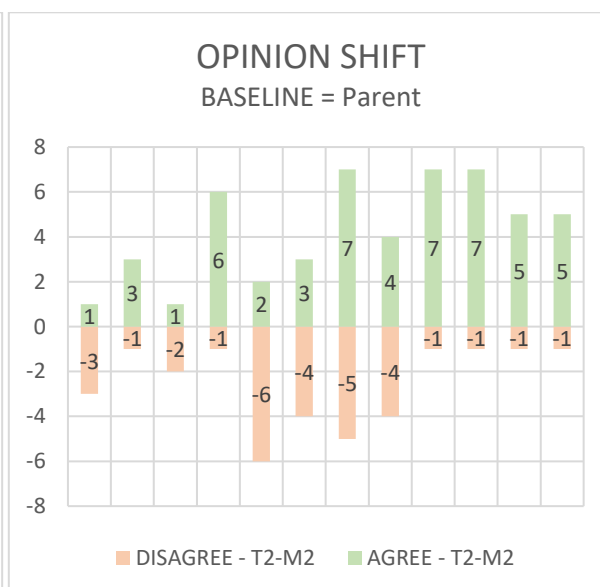
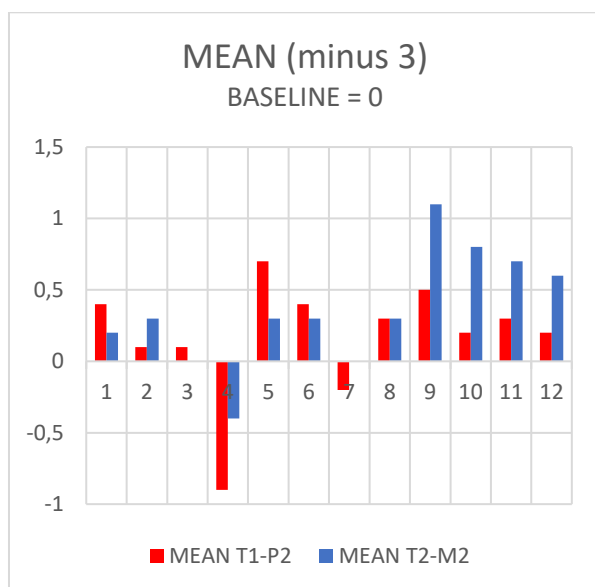
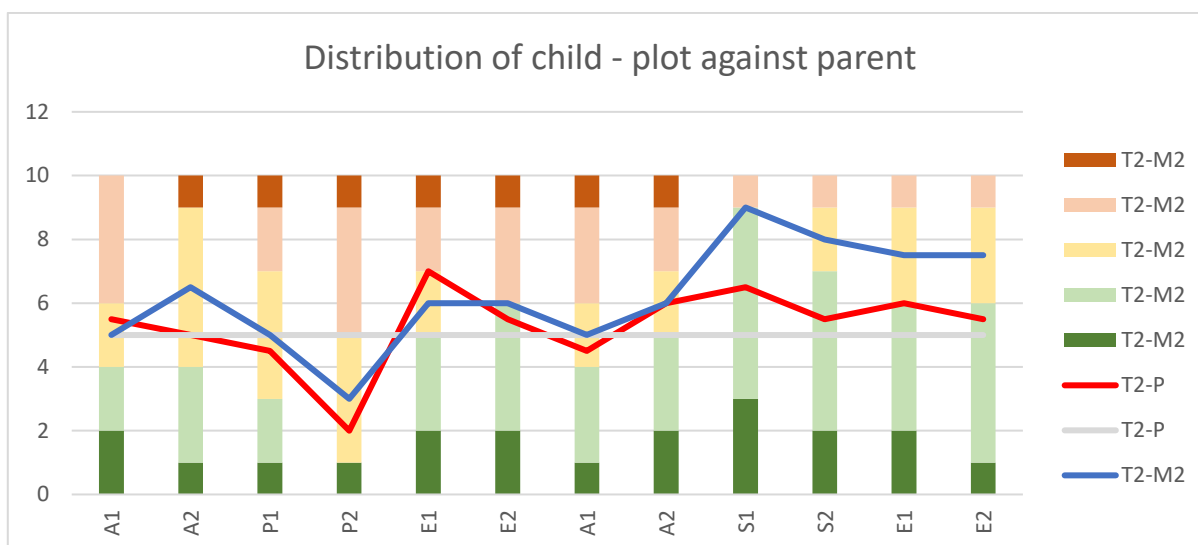


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	3,4	3,2	3,1	2,4	3,8	3,5	3	3,4	3,5	3,2	3,5	3,5

ANALYSES BY GPT		
Facts	Reasoning	Assumptions
1) Small changes can lead to larger improvements over time. 2) Goal setting can positively impact behavior change.	The message focuses on setting a small, achievable goal which encourages the user to take action. This approach aligns with the behavior change technique of goal setting.	1) The recipient is motivated to improve their dietary habits. 2) They are open to setting and working towards goals. 3) Small changes will be effective for the individual.
PARTICIPANT COMMENT		
AP	het moet wel afgestemd zijn op de hulpbehoefte	
AP	good advice, but missing health skills is the reason why people often don't do this	
AP	result goal or behaviour goal? That's a big difference	
AP	haalbaar doel is voor sommige mensen al lastig zelf te interpreteren.	
EA	het moet wel afgestemd zijn op de hulpbehoefte en communicatiemogelijkheden	
EA	You have to know what a small achievable goal is, otherwise people tend to make the goal to big and therefor fail and give up	
EA	It's important how you formulate a goal. On behaviour or on result. Result goal for a week is not effective.	
SE	see above, this statement could set people up for failure if not explained	

CHILD – T2-M2 (Top - 3)

"This week, set a goal to replace one snack with a healthy option! Every change counts."

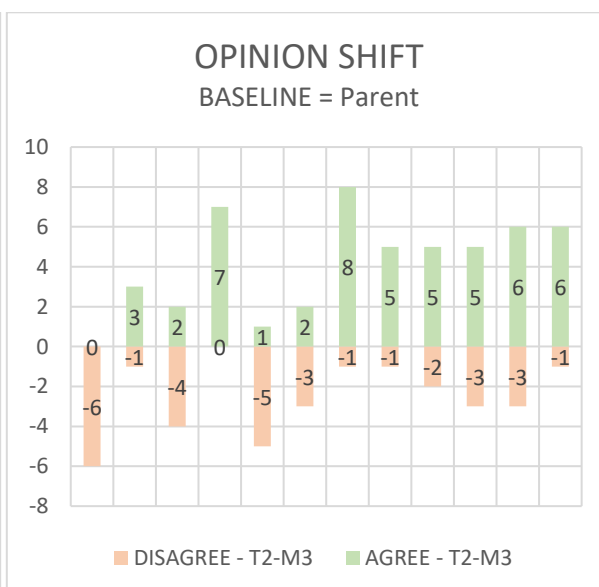
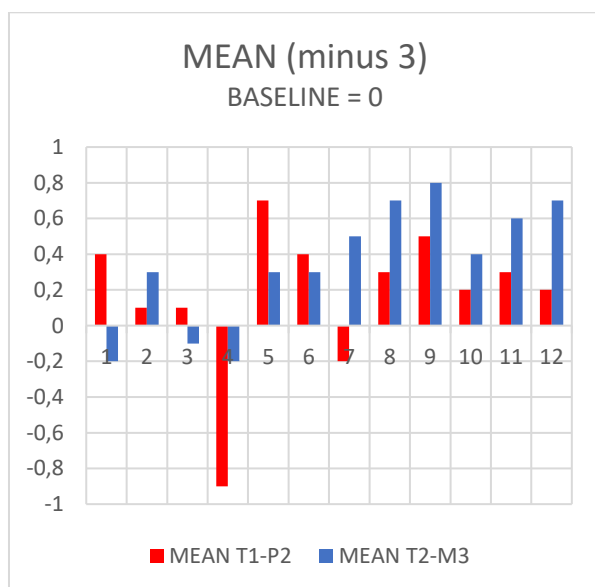
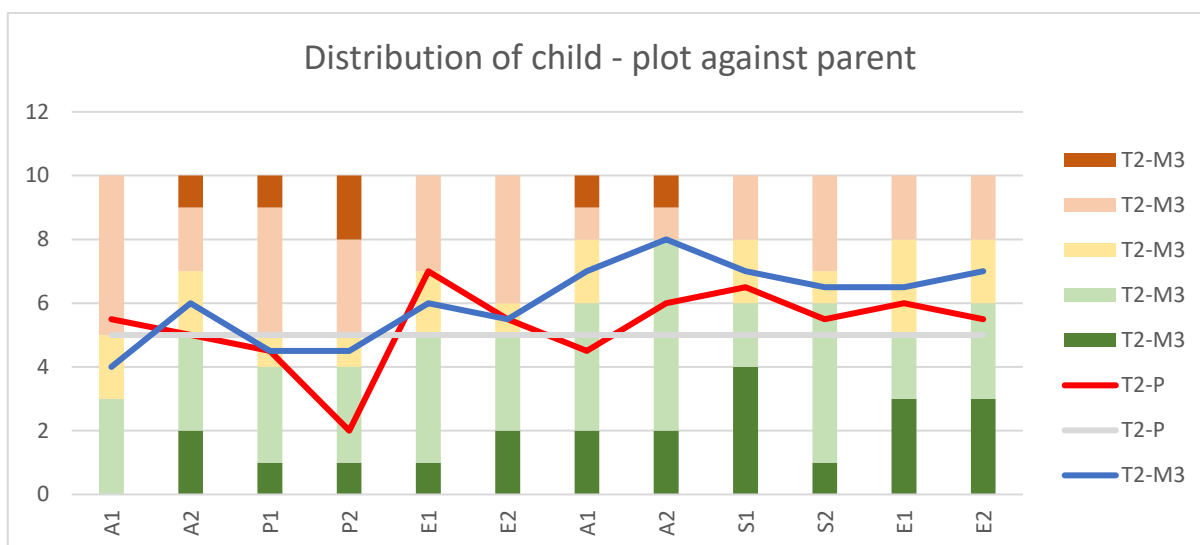


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	3,2	3,3	3	2,6	3,3	3,3	3	3,3	4,1	3,8	3,7	3,6

ANALYSES BY GPT		
Facts	Reasoning	Assumptions
1) Replacing unhealthy snacks can improve overall diet. 2) Goal setting encourages behavior change.	The message suggests a specific, achievable dietary goal to replace a snack with a healthier option, which is in line with the goal setting behavior change technique.	1) The recipient wants to improve their dietary habits. 2) They are open to setting and working towards goals. 3) Replacing a snack can be a meaningful change for the individual.
PARTICIPANT COMMENT		
AP	het moet wel afgestemd zijn op de hulpbehoefte en communicatiemogelijkheden en financiële situatie	
AP	question is what a healthy alternative is, but good advice and stated clear	
AP	In Dutch we say, "een open deur". The healthy snack doesn't manage the mental hunger.	
EA	het moet wel afgestemd zijn op de hulpbehoefte en communicatiemogelijkheden en financiële situatie	
EA	good advice, clear and specific statement	
SA	het moet wel afgestemd zijn op de hulpbehoefte en communicatiemogelijkheden en financiële situatie	

CHILD – T2-M3 (Top - 4)

"Aim for adding one more fruit or veggie to your daily meals. Small goals lead to big results!"

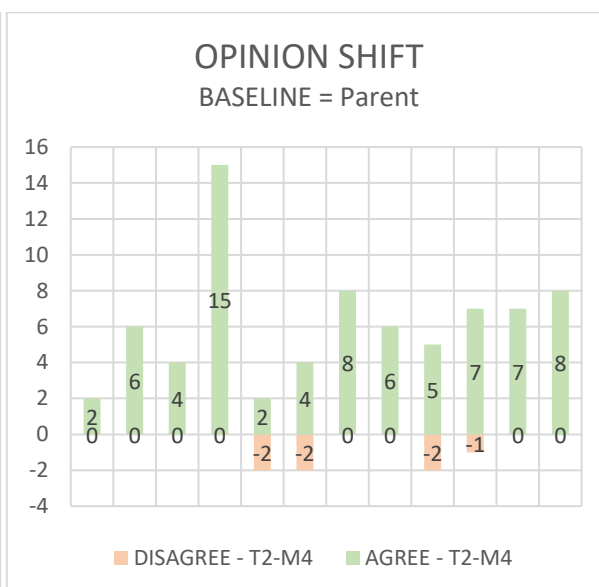
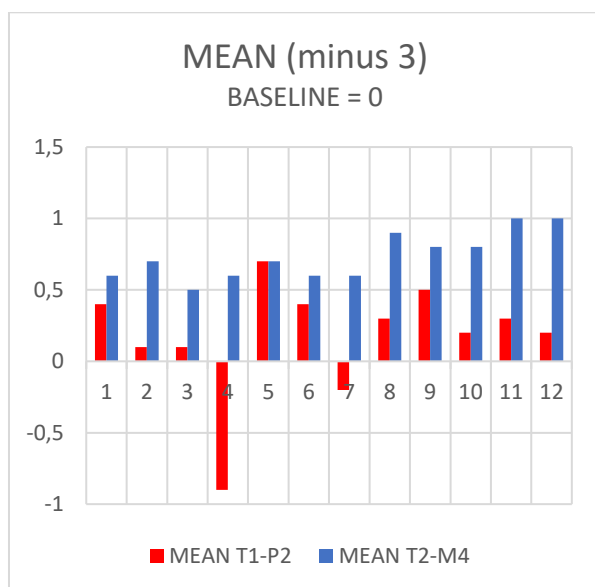
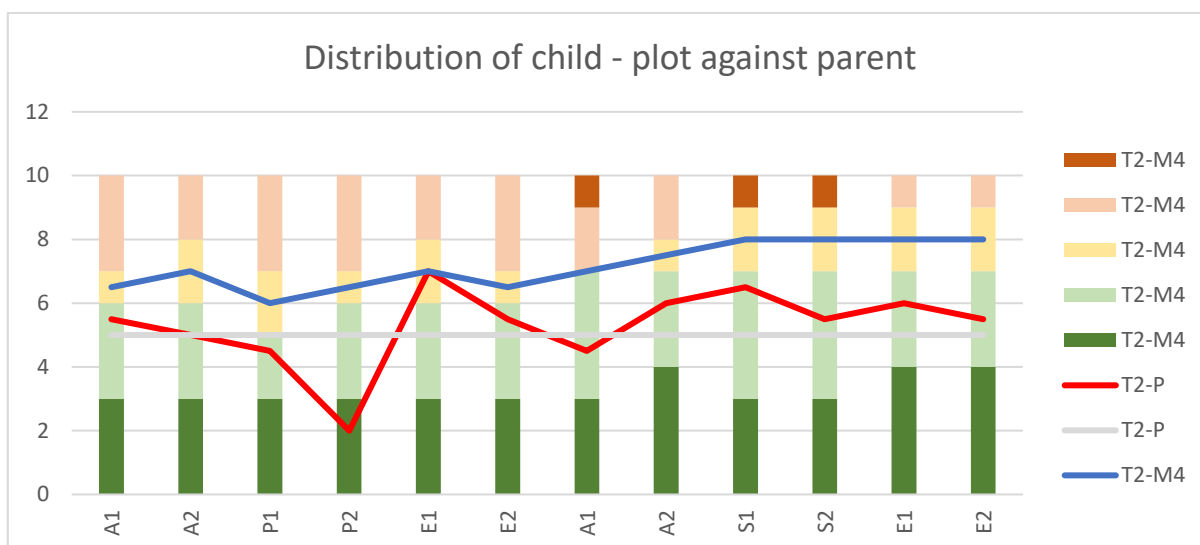


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	2,8	3,3	2,9	2,8	3,3	3,3	3,5	3,7	3,8	3,4	3,6	3,7

ANALYSES BY GPT		
Facts	Reasoning	Assumptions
1) Consuming more fruits and vegetables can improve dietary habits. 2) Goal setting can influence behavior change.	The message promotes a specific, achievable dietary goal of adding a fruit or vegetable to daily meals, which aligns with the goal setting behavior change technique.	1) The recipient is motivated to improve their dietary habits. 2) They are open to setting and working towards goals. 3) Adding a fruit or vegetable will have a positive impact on their diet.
PARTICIPANT COMMENT		
AP	good specific advice and a good statement	
AP	You have to explain veggie and you have to give examples which match with the person.	
SA	Again: "open deur" advice en stimulation	

CHILD – T2-M4 (Top – 2)

"Set a goal to drink more water daily. Small habits make a big difference in the long run!"

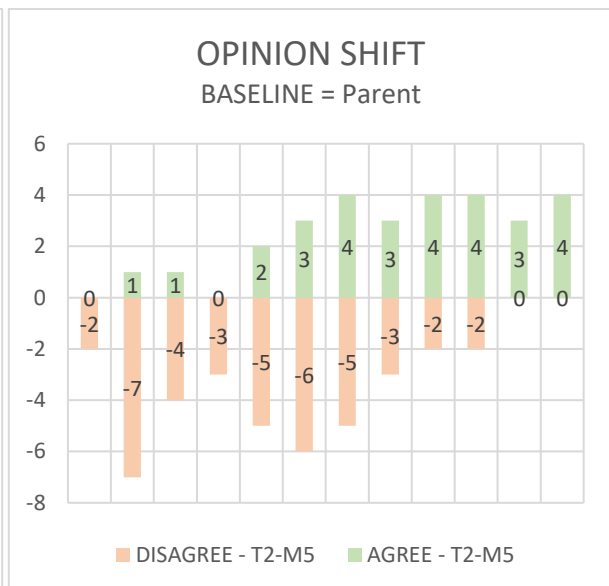
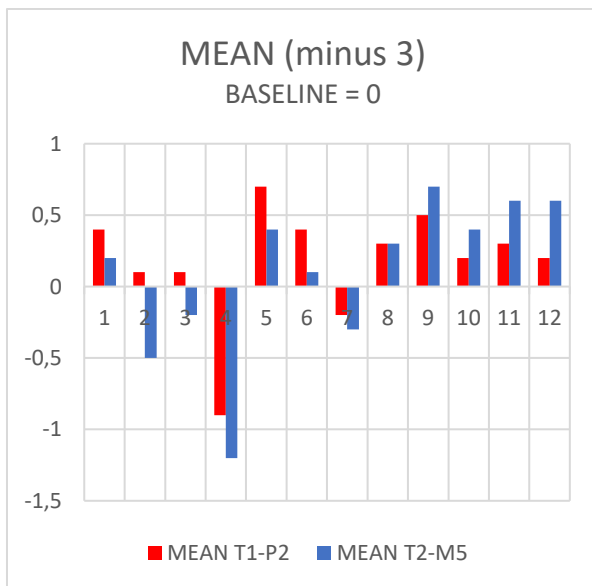
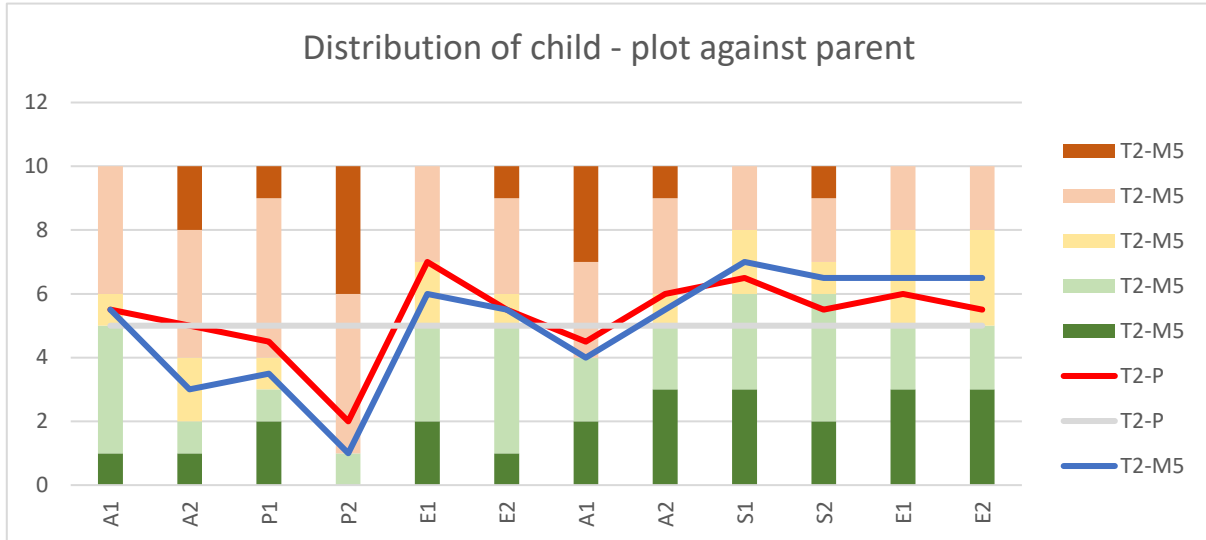


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	3,6	3,7	3,5	3,6	3,7	3,6	3,6	3,9	3,8	3,8	4	4

ANALYSES BY GPT		
Facts	Reasoning	Assumptions
1) Drinking more water can improve overall health. 2) Goal setting can lead to positive behavior change.	The message suggests a specific, achievable goal of increasing daily water intake, which is consistent with the goal setting behavior change technique.	1) The recipient wants to improve their dietary habits. 2) They are open to setting and working towards goals. 3) Drinking more water will have a positive impact on their health.
PARTICIPANT COMMENT		
AP	instead of what? or just more? better if it was more specific	
AP	It's about behaviour, it's not one-way traffic information and advice	

CHILD – T2-M5

"Try to reduce added sugars this week. Setting small goals can pave the way for lasting change!"

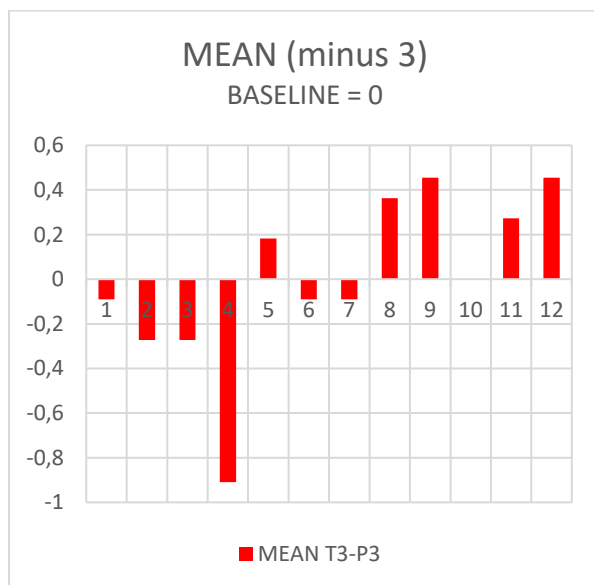
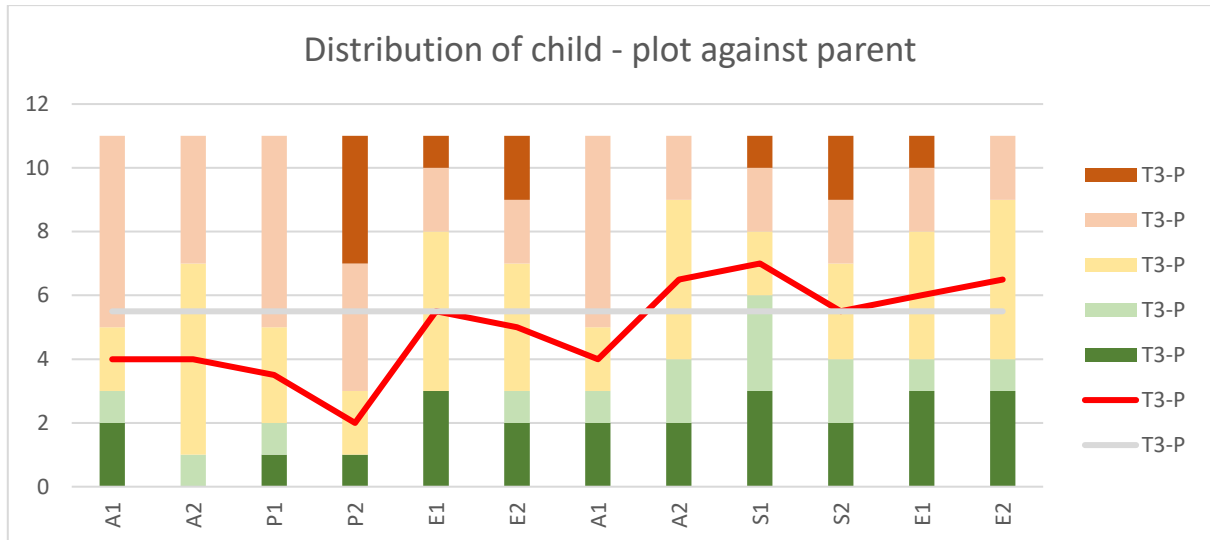


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	3,2	2,5	2,8	1,8	3,4	3,1	2,7	3,3	3,7	3,4	3,6	3,6

ANALYSES BY GPT		
Facts	Reasoning	Assumptions
1) Reducing added sugars can improve dietary habits. 2) Goal setting is effective for behavior change.	The message encourages setting a specific, achievable goal to reduce added sugars, which aligns with the goal setting behavior change technique.	1) The recipient is motivated to improve their dietary habits. 2) They are open to setting and working towards goals. 3) Reducing added sugars will be beneficial for the individual.
PARTICIPANT COMMENT		
AP	what is added sugar? how do you recognize that? knowledge about that isnt common. Better would be "try to avoid beverages with added sugar"	
AP	niet iedereen weet waar toegevoegde suikers in zitten (in de producten die die persoon reeds gebruikt)	
EA	unclear what people think what added sugar is, should make this more specific	

PARENT – T3-P

"Each day, reward another step towards your goals. Today, reward yourself for eating healthy. Tomorrow, reward yourself for cooking that healthy meal!"

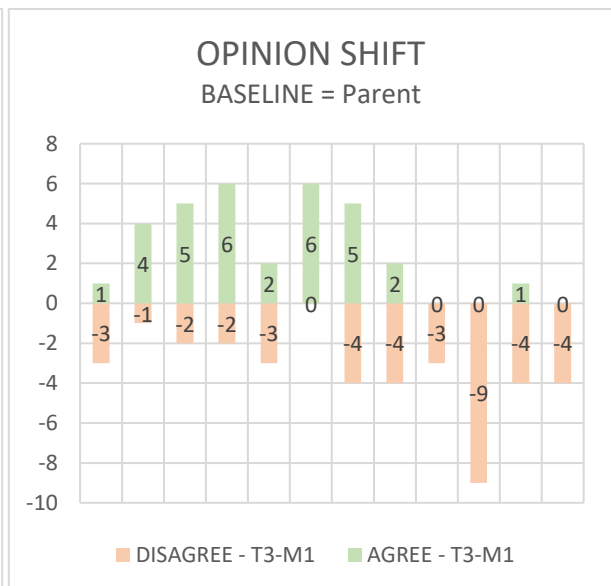
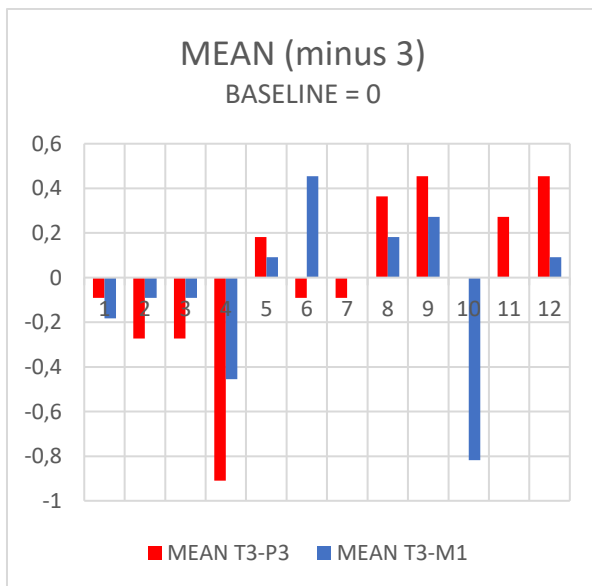
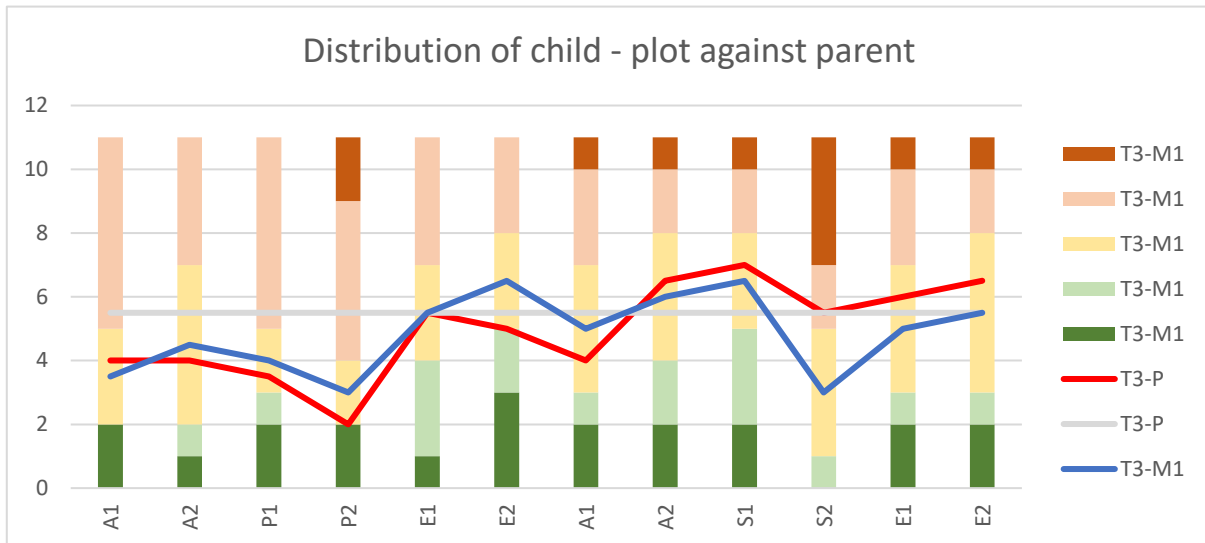


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	2,9	2,7	2,7	2,1	3,2	2,9	2,9	3,4	3,5	3,0	3,3	3,5

PARTICIPANT COMMENT	
EA	Belonen werkt niet bij duurzame gedragsverandering. Als Motivatie van binnenuit versterkt wordt wel. Is veelal maatwerk
EA	ligt aan de soort beloning men kiest
EA	Ingewikkelde opdracht, vraagt te veel mijn inziens
SA	te weinig concreet
SA	best veel tekst voor laaggeletterden
SA	Het is belangrijk dat de berichten goed aansluiten op wat door de diëtist al is uitgelegd. Als dat zo is dan is dit een stimulans om door te gaan.

CHILD – T3-M1

"Celebrate progress with a self-reward for each healthy choice made. You deserve it!"

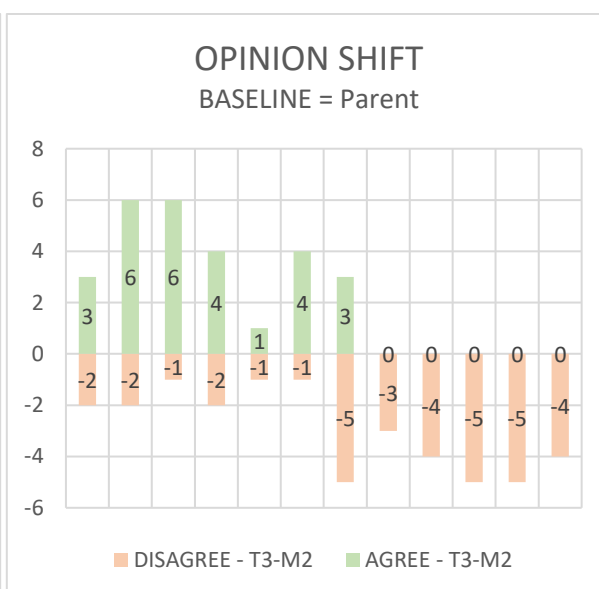
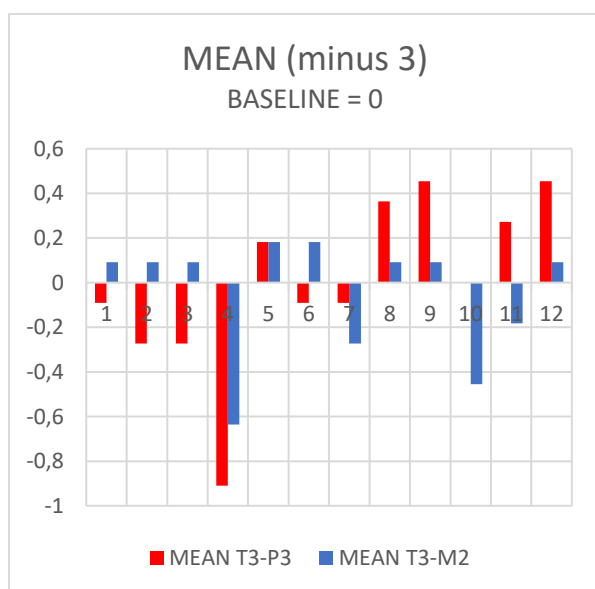
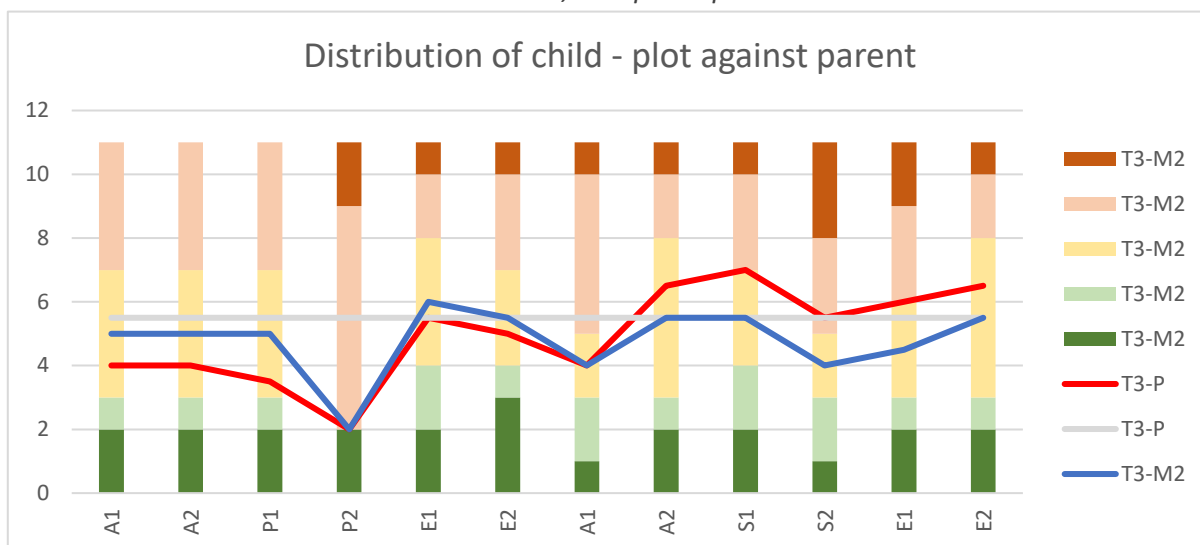


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	2,8	2,9	2,9	2,5	3,1	3,5	3,0	3,2	3,3	2,2	3,0	3,1

ANALYSES BY GPT		
Facts	Reasoning	Assumptions
1) Self-reward is effective in reinforcing healthy behaviors 2) Progress has been made in improving dietary habits 3) Rewards can be non-food items or experiences.	Using the self-reward behavior change technique, we encourage individuals to recognize their progress and reward themselves. This positive reinforcement helps build a stronger connection between healthy habits and positive emotions.	1) The recipient is making progress towards better dietary habits 2) They respond positively to self-reward as a motivator 3) They can identify suitable non-food rewards for themselves.
PARTICIPANT COMMENT		
AP	Externe beloning werkt maar tijdelijk, niet zo zinvol;	
AP	Deze bemoediging kan volgen na een uitleg en inventarisatie hoe iemand zichzelf kan belonen. Mensen denken toch heel snel in gekochte beloningen, met name ook voor kinderen. Mensen met weinig geld menen dat ze een flinke beloning in geld/waarde moeten geven.	
EA	Zie mijn eerste opmerking. Mensen uit een andere cultuur zul je echt moeten helpen met mogelijkheden voor beloning voor zichzelf. (DIRECT HIERBOVEN)	
SA	welke beloning wordt gekozen heeft effect en vereist goede afweging	
SA	haha dit gaat geweldige resultaten opleveren ; gebakje voor elk stuk fruit	

CHILD – T3-M2 (Low - 4)

"Acknowledge your healthy choices today and treat yourself! Every step counts, keep it up!"

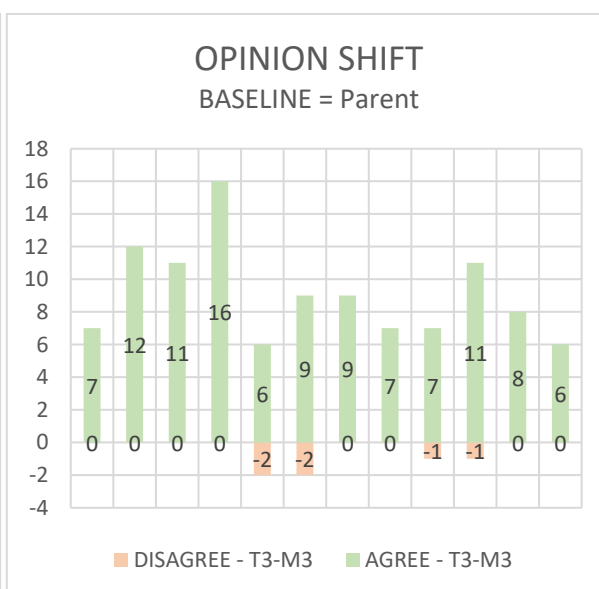
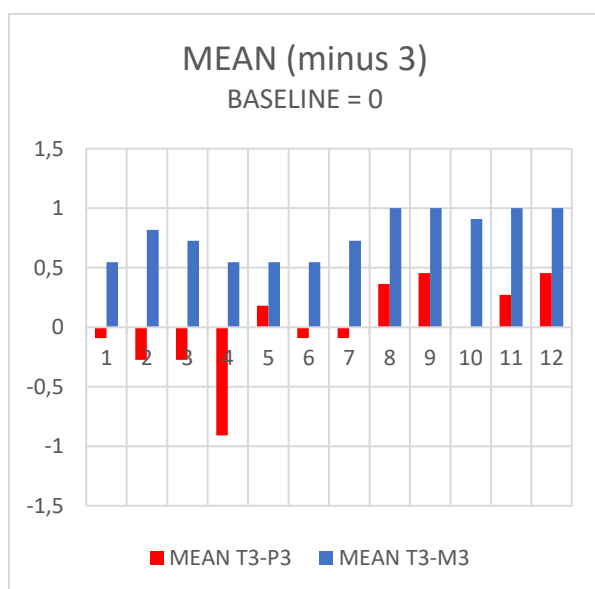
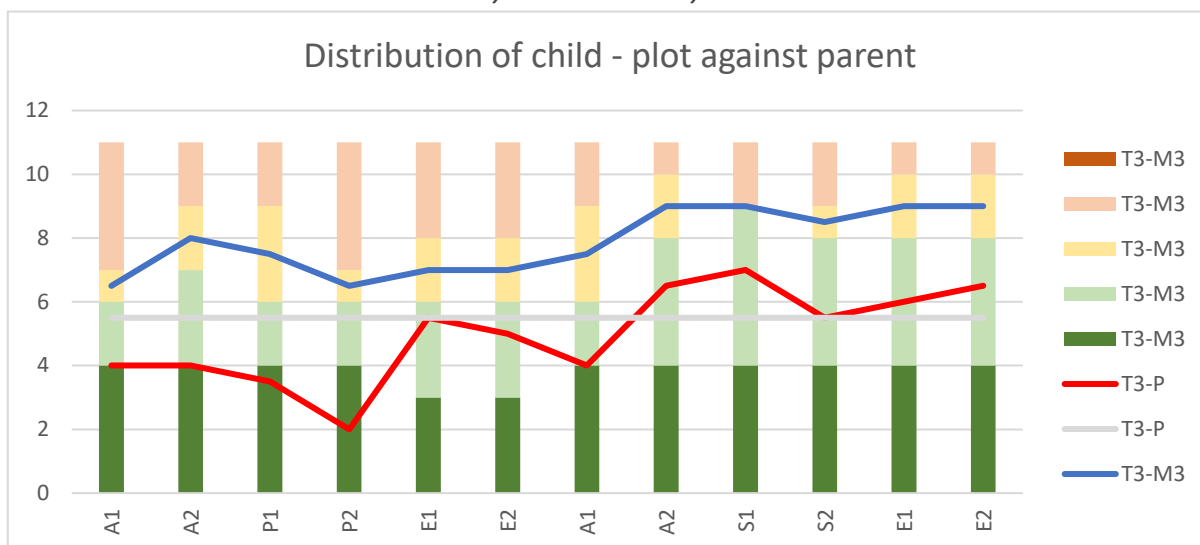


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	3,1	3,1	3,1	2,4	3,2	3,2	2,7	3,1	3,1	2,5	2,8	3,1

ANALYSES BY GPT		
Facts	Reasoning	Assumptions
1) Self-reward is effective in reinforcing healthy behaviors 2) Progress has been made in improving dietary habits 3) Rewards can be non-food items or experiences.	Using the self-reward behavior change technique, we promote recognizing and celebrating every small achievement in the journey towards better dietary habits. This helps to reinforce positive behaviors and maintain motivation.	1) The recipient is making progress towards better dietary habits 2) They respond positively to self-reward as a motivator 3) They can identify suitable non-food rewards for themselves.
PARTICIPANT COMMENT		
AP	Maakt veel uit wat als 'treat' wordt gekozen	
EA	Wie weet wat een gezonde keuze is? Er is veel verwarring bij alg publiek. En welke beloning wordt gekozen?? Is niet genoeg motiverend voor duurzame gedfdragsverandering	
EA	ervan uitgaande dat de betreffende beloning ook gezond is	

CHILD – T3-M3 (Top – 1)

"Pat yourself on the back for each healthy choice you make. You're on your way to a better you!"

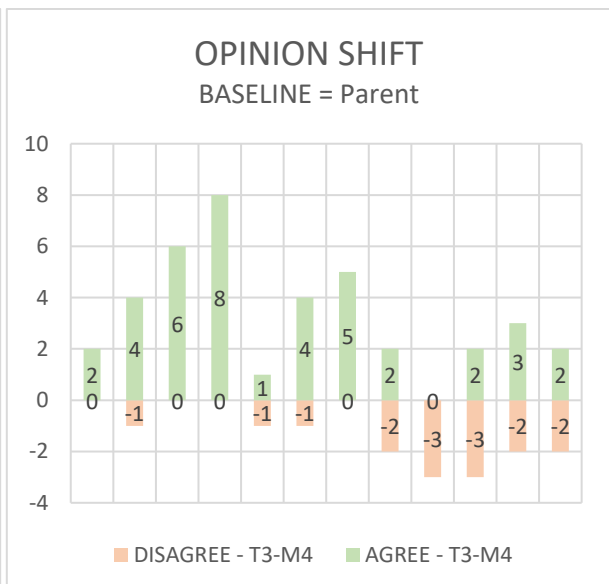
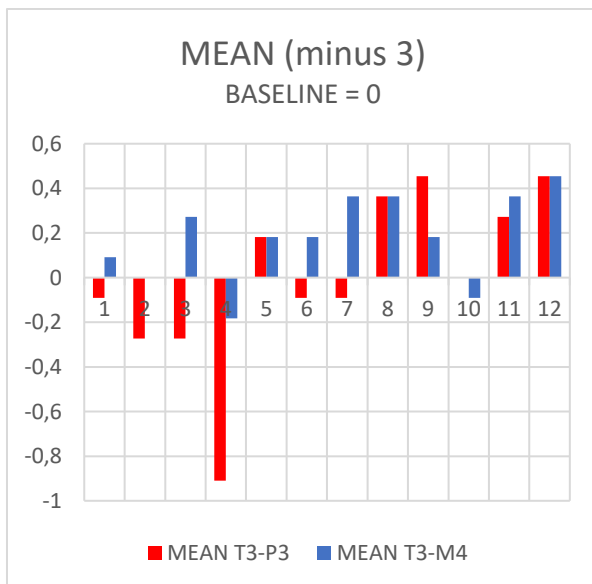
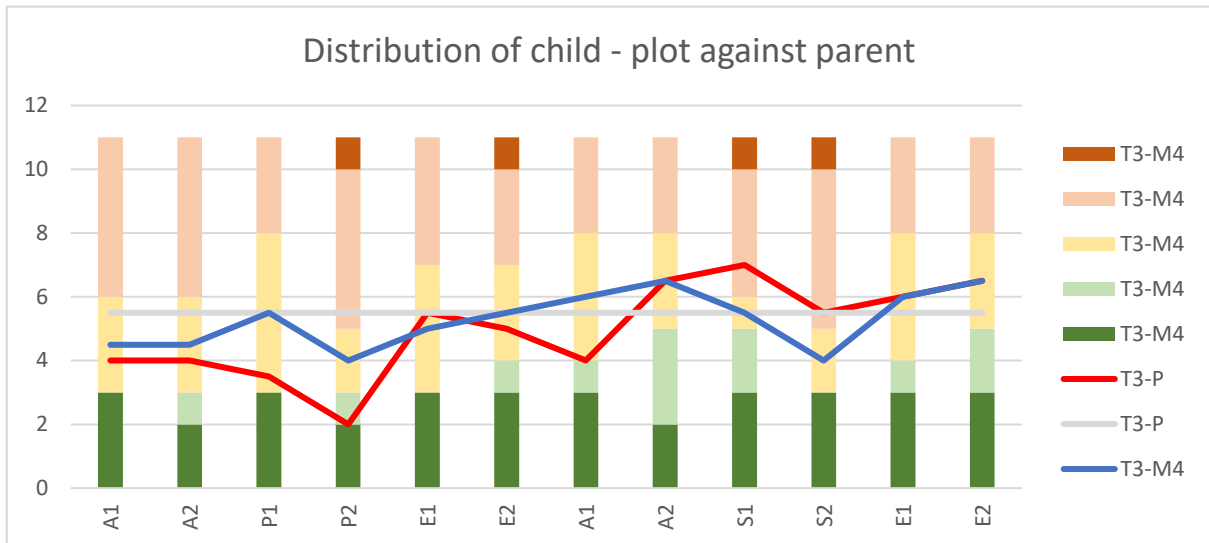


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	3,5	3,8	3,7	3,5	3,5	3,5	3,7	4,0	4,0	3,9	4,0	4,0

ANALYSES BY GPT		
Facts	Reasoning	Assumptions
1) Self-reward is effective in reinforcing healthy behaviors 2) Progress has been made in improving dietary habits 3) Rewards can be non-food items or experiences.	The message uses the self-reward behavior change technique by encouraging individuals to praise themselves for their healthy choices. This helps to build motivation and maintain healthy habits.	1) The recipient is making progress towards better dietary habits 2) They respond positively to self-reward as a motivator 3) They can identify suitable non-food rewards for themselves.
PARTICIPANT COMMENT		
AP	eerst helpen healthy choices helder formuleren, op maat per persoon	

CHILD – T3-M4

"Savor your victories in healthy eating. Reward yourself and be proud of the strides you're making!"

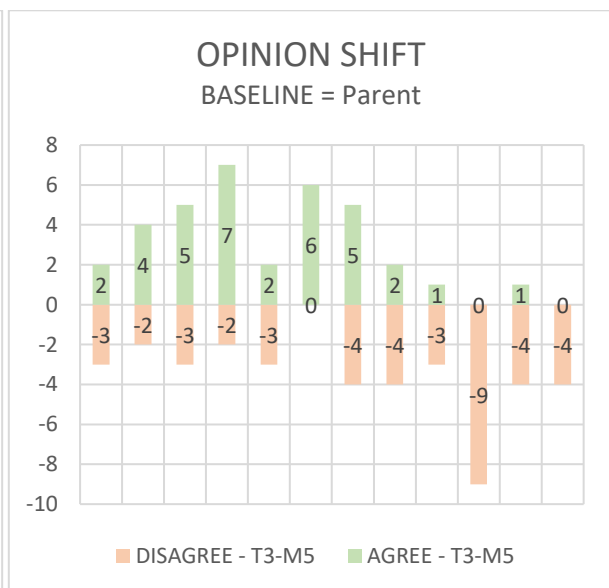
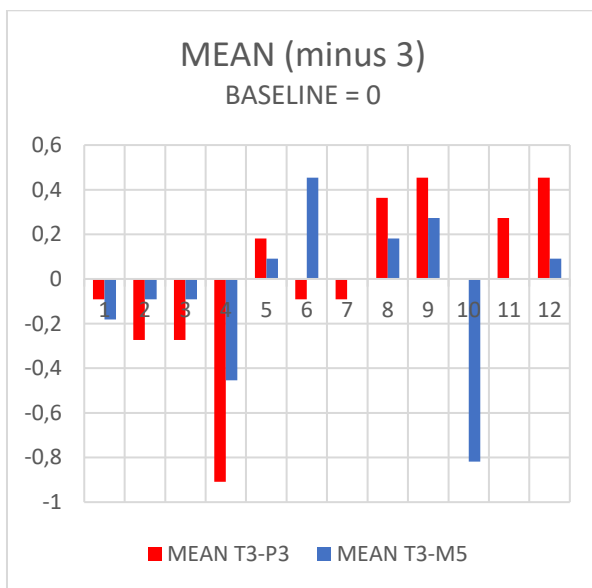
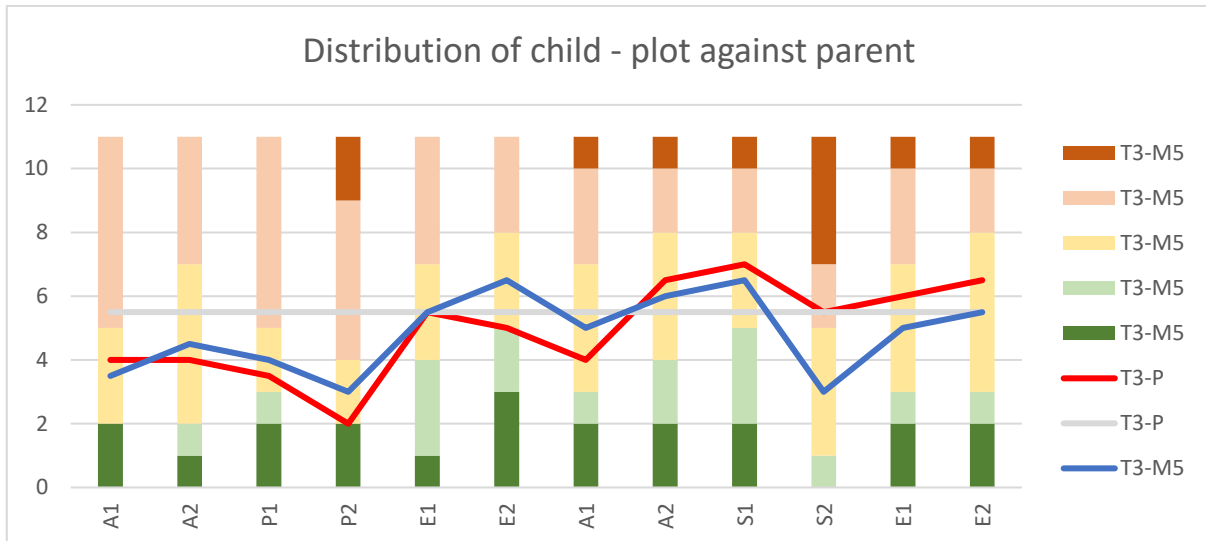


	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	3,1	3,0	3,3	2,8	3,2	3,2	3,4	3,4	3,2	2,9	3,4	3,5

ANALYSES BY GPT		
Facts	Reasoning	Assumptions
1) Self-reward is effective in reinforcing healthy behaviors 2) Progress has been made in improving dietary habits 3) Rewards can be non-food items or experiences.	The message employs the self-reward behavior change technique, emphasizing the importance of recognizing and celebrating achievements in healthy eating. This helps to keep motivation high and reinforce positive habits.	1) The recipient is making progress towards better dietary habits 2) They respond positively to self-reward as a motivator 3) They can identify suitable non-food rewards for themselves.
PARTICIPANT COMMENT		
AP	Als alles beloond moet worden, krijg je geen gedragsverandering. De beloning zit in de eigen regie nemen en de innerlijke drijfveren, vaak ook op lange termijn beloning	

CHILD – T3-M5 (Low - 5)

"Every healthy choice is a win! Reward yourself for progress made and stay committed to your goals!"



	A1	A2	P1	P2	E1	E2	A1	A2	S1	S2	E1	E2
MEAN	2,8	2,9	2,9	2,5	3,1	3,5	3,0	3,2	3,3	2,2	3,0	3,1

ANALYSES BY GPT		
Facts	Reasoning	Assumptions
1) Self-reward is effective in reinforcing healthy behaviors 2) Progress has been made in improving dietary habits 3) Rewards can be non-food items or experiences.	The message uses the self-reward behavior change technique to encourage rewarding oneself for making healthy choices. This helps to maintain motivation and reinforce the connection between healthy habits and positive emotions.	1) The recipient is making progress towards better dietary habits 2) They respond positively to self-reward as a motivator 3) They can identify suitable non-food rewards for themselves.
PARTICIPANT COMMENT		
Effect hangt ook af van de soort beloning die mensen zich dan gunnen		

A8.3 Histograms for Assessing data distribution

A sample of the data was taken to visualize histograms to analyse the data distribution.

