## Empowering vulnerable people with serious games and gamification

van der Lubbe, Laura Marceline

2022

***document version***
Publisher's PDF, also known as Version of record

# Empowering Vulnerable People with Serious Games and Gamification

Laura van der Lubbe

# Empowering vulnerable people with serious games and gamification

L.M. van der Lubbe

VRIJE UNIVERSITEIT AMSTERDAM

SIKS Dissertation Series No. 2022-17
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

De omslag van dit proefschrift is ontworpen door Ruth Govaerts.

Geprint door: Print Service Ede - www.proefschriftenprinten.nl

ISBN/EAN: 978-90-832468-6-4

VRIJE UNIVERSITEIT

# Empowering vulnerable people with serious games and gamification

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Bètawetenschappen
op woensdag 29 juni 2022 om 13.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

Laura Marceline van der Lubbe

geboren te Amstelveen

| | |
|---|---|
| promotor: | prof.dr. K.V. Hindriks |
| copromotoren: | dr.mr. C. Gerritsen |
| | dr. M.C.A. Klein |
| promotiecommissie: | prof.dr. M. Hoogendoorn |
| | prof.dr. M.M. Riper |
| | dr.ing. S.C.J. Bakkes |
| | prof.dr.ir. P. Spronck |
| | dr. W.P. Brinkman |

Eat, drink, be happy.

*Eet, drink, wees gelukkig.*

# Summary

Although many people will associate games with entertainment and leisure, games can also aim more serious purposes, such as training or education. Games with such goals are called *serious games*. In addition, *gamification* means that a (serious) task is enhanced with game elements. Examples of serious games and gamification can be seen in many aspects of daily life. Loyalty programs of stores, educational games in schools, fitness wearables and their gamified applications, rehabilitation games, and so on. In this dissertation, the focus is on a specific domain in which serious games and gamification can create societal benefit, namely by using them to empower vulnerable target groups.

In the first part of this dissertation, a literature review is performed to understand the domain of serious games and gamification for vulnerable target groups. Based on this review, research gaps can be identified. Moreover, the review resulted in a taxonomy that is used throughout the dissertation to classify different games and applications.

In the following parts of the dissertation, projects addressing two different target groups and in total three vulnerabilities are discussed. The first target group is older adults, who are vulnerable in different ways. In this dissertation, safety risks for doorstep scams and health risks through malnutrition are addressed. The first vulnerability is addressed by a serious game using interactive scenarios of doorstep scams. A diet tracking system that was used to support participants in a diet trial addressed the latter vulnerability. The second target group is young adults, which is an age group with a vulnerable mental well-being. The last part of this dissertation aims to study how gamification can be used to enhance self-compassion among young adults via an online 6-weeks training program, to increase their resilience in the face of mental well-being difficulties.

Artificial Intelligence (AI) technologies can be used to personalize and adapt the experience of a game to users. Tone of voice analysis was used to influence the progression in scenarios of the serious game about doorstep scams, and it gave players the possibility to assess the assertiveness of their voice. Machine learning algorithms were used to create personalized meal recommendations that can be used to improve the user experience of the diet tracking system for older adults. These algorithms base their recommendations on information about the historical intake of users to suggest meals and to additional items during meal editing. This makes the process of registering a meal less time-consuming. Sentiment analysis is used to adapt responses of the system in an exercise from the self-compassion training program. In addition, a topic detection algorithm was designed to assign one topic from a predefined set of topics to a note by a user of the training program. With this information, users can choose different types of situations to use in the exercises: frequently or rarely discussed topics. Aside from those techniques, knowledge representation is used in all projects, which is important for serious games/gamified applications since they are often based on expert and/or domain knowledge.

This dissertation contributes to understanding the domain of serious games and gamification to empower vulnerable groups. The work also contributes to the research on the development of applications within that domain. On top of that, it contributes to understanding how AI techniques can be used to offer (personalized) features that enrich serious games or gamified applications. Finally, for each of the project centered parts, the results that are found in those parts contribute to the research in those specific fields.

# Samenvatting

Veel mensen zullen games associëren met entertainment en ontspanning, maar games kunnen ook serieuzere doelen hebben zoals training of onderwijs, ze worden dan serious games genoemd. Daarnaast is gamification het toevoegen van spelelementen aan een (serieuze) taak. Voorbeelden van serious games en gamification zijn overal te vinden in het dagelijkse leven. Spaarprogramma's van winkels, educatieve games op scholen, fitnesshorloges en de bijbehorende applicaties, revalidatiegames, enzovoorts. In dit proefschrift ligt de focus op een specifiek domein waarbij serious games en gamification een maatschappelijk doel hebben, namelijk om kwetsbare groepen te emanciperen.

Voor het eerste deel van dit proefschrift is een literatuurreview uitgevoerd om dit specifieke domein te begrijpen. Op basis van deze review kunnen hiaten in het bestaande onderzoek worden geïdentificeerd en is een taxonomie gemaakt die in de rest van dit proefschrift wordt gebruikt om verschillende games en applicaties te classificeren.

In de volgende delen worden projecten voor twee doelgroepen en in totaal drie kwetsbaarheden besproken. De eerste doelgroep is ouderen, die om verschillende redenen kwetsbaar kunnen zijn. Dit proefschrift onderzoekt veiligheidsrisico's door babbeltrucs en gezondheidsrisico's van ondervoeding. De eerste kwetsbaarheid wordt aangepakt met een serious game met interactieve babbeltrucscenario's. Een dieetregistratiesysteem om deelnemers in een dieetstudie te ondersteunen richt zich op de andere kwetsbaarheid. De tweede doelgroep is jongvolwassenen: een leeftijdsgroep met een kwetsbaar mentaal welzijn. In het laatste deel wordt gekeken hoe gamification kan worden gebruikt om zelfcompassie bij deze doelgroep te vergroten met een 6 weken durende online training om de weerbaarheid tegen moeilijkheden voor hun mentale welzijn te vergroten.

Het gebruik van kunstmatige Intelligentie (KI) technieken kan de gebruikerservaring met een game personaliseren of aanpassen. Stemtoonanalyse werd gebruikt om het verloop van de babbeltrucscenario's te beïnvloeden en spelers de assertiviteit van hun stem te laten beoordelen. Machine learning algoritmes werden gebruikt om persoonlijke aanbevelingen voor maaltijden te geven, om daarmee de gebruikerservaring met het dieetregistratiesysteem te verbeteren. Deze algoritmes gebruiken informatie over de eetgeschiedenis van gebruikers, om zo maaltijden of extra producten tijdens het aanpassen van een maaltijd aan te bevelen. Dit kan maaltijdregistratie minder tijdrovend maken. Sentimentanalyse werd gebruikt om de reacties van het systeem voor één van de oefeningen van de zelfcompassietraining aan te passen. Daarnaast is er een onderwerp-detectiealgoritme ontwikkeld om het onderwerp van een tekst van een gebruiker te kiezen uit een vastgestelde lijst. Door deze informatie kunnen gebruikers kiezen uit verschillende soorten situaties voor oefeningen: vaak of zelden besproken onderwerpen. Naast deze technieken is kennisrepresentatie in alle projecten gebruikt. Dit is belangrijk voor serious games en gamification omdat deze vaak gebaseerd zijn op expert- en/of domeinkennis.

Dit proefschrift draagt bij aan het begrijpen van het domein van serious games en gamification om kwetsbare groepen te emanciperen. Het werk draagt ook bij aan onderzoek naar de ontwikkeling van toepassingen in dat domein. Daarnaast draagt het bij aan het begrip hoe KI-technieken (gepersonaliseerde) functies kunnen bieden ter verrijking van serious games of gamification. Tot slot dragen de resultaten van de projecten van dit proefschrift bij aan het onderzoek in die specifieke velden.

# Contents

# Dankwoord

1 september 2012: De VU opent weer een academisch jaar. Voor mij is het mijn allereerste academische jaar. Als jonge studente liep ik zonder al te veel verwachtingen een nieuw avontuur tegemoet. Ik had geen duidelijk toekomstplan en had nog geen idee dat ik bijna 10 jaar bij de VU zou blijven hangen. In die 10 jaar heb ik me kunnen ontwikkelen van jonge studente tot de Doctor die ik na de verdediging van dit proefschrift hopelijk kan zijn. Die ontwikkeling gaat verder dan alleen een vakinhoudelijke ontwikkeling. Ik heb tijdens mijn studie en promotietraject ook ontdekt waar mijn interesses en kwaliteiten nog meer liggen. Zo ontdekte ik dat ik me wil en moet inzetten voor meer vrouwen in (bèta)-wetenschap en beroepen, en dat ik hou van informatief en wetenschappelijk schrijven. Maar ook ontdekte ik stukjes van de wereld, tijdens vakanties of reizen voor conferenties en door nieuwe mensen te ontmoeten, met andere achtergronden, culturen, en overtuigingen.

Dat ik me als persoon de afgelopen 10 jaar heb kunnen ontwikkelen, is ook te danken aan anderen die ik tijdens deze periode heb ontmoet, die me gesteund hebben, of waarmee ik waardevolle herinneringen heb gemaakt. In dit dankwoord wil ik de belangrijkste mensen in het zonnetje zetten; op de manier die het beste bij mij past: in geschreven vorm. Daarbij wilde ik in dit dankwoord ook citaten opnemen van inspirerende vrouwen, om mijn verhaal te ondersteunen, of om een boodschap te delen in woorden die ik zelf niet kon vinden.

> "A woman with a voice is by definition a strong woman. But the search to find that voice can be remarkably difficult."
> — Melinda Gates

Pas toen ik mij ging verdiepen in de problematiek rondom de genderdisbalans in technische studies en beroepen, realiseerde ik me dat ik zelf kansen heb gekregen die andere meisjes vaak niet krijgen. Meisjes die ten onrechte te horen krijgen van ouders of docenten dat ze ‘niet goed zijn in wiskunde’ of ‘een studie met computers niets voor hen is’. Pap en mam, dank jullie wel dat jullie mij nooit het gevoel hebben gegeven dat ik iets niet kon omdat ik een meisje was. Zelfs anno 2022 geloven niet alle ouders dat. Bedankt dat jullie me zorgeloos lieten studeren, en voor alle steun.

> “What if, in raising children, we focus on ability instead of gender? What if we focus on interest instead of gender?”
> — Chimamanda Ngozi Adichie, We Should All Be Feminists

Lieve Stefan, bedankt voor al je hulp, steun, gezelligheid en liefde. Hoewel dit natuurlijk mijn prestatie is ben ik blij dat ik jou daarbij aan mijn zijde heb, als paranimf, als steun en toeverlaat, als verloofde. Je bent een enorme steun en hulp geweest, door papers te lezen of te luisteren naar presentaties die ik wilde oefenen. In de afgelopen 2 jaar waren we ineens ook thuiswerkcollega's. Jij weet daardoor als geen ander welke ups en downs deze PhD heeft gehad. Dank voor al je steun in de downs, maar ook zeker tijdens de ups. Er zijn eigenlijk geen woorden om te beschrijven wat jouw onvoorwaardelijke trots en steun voor mij betekent, dus hou ik het bij de woorden die in jouw trouwring komen te staan: $\sigma'\alpha\gamma\alpha\pi\acute{\omega}$.

Een vangnet is voor mij meer dan alleen familie. Daarom een hele dikke dank je wel aan alle vriendinnen die in de afgelopen jaren een steun of afleiding voor me waren. Door samen iets leuks te doen, praktisch ergens mee te helpen of gewoon te luisteren als ik mijn hart wilde luchten. Jullie weten wie jullie zijn.

> "True friends are families which you can select."
> — Audrey Hepburn

In het bijzonder wil ik Romy en Gossa bedanken. We hebben in de afgelopen jaren al zoveel met elkaar gedeeld, tijdens onze studie, als vriendinnen en in Romy's geval ook als collega's. Onze vriendschappen betekenen enorm veel voor mij. Ik kan altijd op jullie rekenen; of dat nou is voor emotionele steun, een gezellig avondje reality tv of praktische hulp bij wat dan ook. Dank jullie wel! Ik weet zeker dat we nog heel veel nieuwe, dierbare, herinneringen zullen maken samen.

De omslag van dit proefschrift is ontworpen door Ruth Govaerts. We kennen elkaar al zo'n 12 jaar en hebben vele uurtjes doorgebracht in onze "stamkroeg" en op vele terrasjes. We gaan 'way back' en ik ben blij dat uit het oog bij ons nooit uit het hart is geweest. Ik vind het heel bijzonder dat jij als professionele illustrator mijn proefschriftomslag hebt kunnen ontwerpen. Én dan gaan we deze zomer ook allebei trouwen! Geweldig dat we deze belangrijke momenten nog steeds samen beleven! Bedankt ook aan jouw verloofde Arteau die zeer charmant heeft geposeerd voor de schaduw op de omslag.

Helaas kan ik deze belangrijke mijlpaal niet met iedereen delen. In het bijzonder ontbreekt er een opvallende persoonlijkheid: mijn middelbare schoolvriendin Saskia die aan het begin van 2016 overleed. Lieve Sas, wat had ik je er graag bij gehad tijdens mijn verdediging. Ik weet zeker dat je m'n grootste cheerleader was geweest. Ik geniet daarom ietsje meer van dat soort momenten, toost een glaasje extra, en glim nog iets meer van trots; speciaal voor jou.

> "Pour être irremplaçable, il faut être different."
> — Coco Chanel

De weg naar mijn PhD was niet de standaard route, en ik ben dan ook een aantal (oud)collega's dankbaar voor hun inzet om dit voor mij mogelijk te maken. Charlotte en Tibor, bedankt dat jullie meedachten over de mogelijkheden vanaf het moment dat ik Junior Researcher was. Michel, bedankt dat ook jij mogelijkheden zag toen ik informeerde naar de positie bij PROMISS. Ik ken jullie al vanuit mijn studie, en heb vanaf dat moment veel van jullie geleerd. Niet alleen vakinhoudelijk, maar ook over het schrijven van papers en over de academische wereld; daarbij was Charlotte voor mij een van de (schaarse) vrouwelijke rolmodellen. Dat was heel belangrijk voor me, dank je wel daarvoor.

> "Success isn't about how much money you make, its about the difference you make in peoples lives."
> — Michelle Obama

Naast de rol die Charlotte en Michel speelden in het vinden van een mogelijkheid om van mijn werk aan de VU ook een PhD te maken, hebben zij samen met Koen ook mijn PhD begeleid. Ik heb veel geleerd van jullie in de afgelopen jaren, een fundament waar ik nog lang op zal kunnen doorbouwen. Ik wilde daarnaast alles op alles zetten om dit proefschrift nog voor de zomer van 2022 af te ronden. Dank voor jullie hulp bij het realiseren van deze droom. Hiervoor ben ik ook dank verschuldigd aan de leescommissie.

> "When you have a dream, you've got to grab it and never let go."
> — Carol Burnett

Also a word of thanks to my Social AI colleagues, who might not speak Dutch: I have very warm memories about the time we spent in the W&N-building where I have shared different offices with different colleagues. When we moved to the NU-building, we all become office buddies in the 'office garden'. I had a lot of fun during my time at the VU, also because of you. Thank you! I also want to thank all the colleagues outside my research group that I collaborated with in some way. Because my research was multidisciplinary, I learnt a lot about other, non-AI, topics, and also about other working and research methods. That was a very instructive experience that helped me develop as a researcher. A special thanks to Mojca, who was always happy to help when needed and did so with great enthusiasm.

Tot slot wil ik nog een aantal andere mensen bedanken voor de inspiratie die ze mij, op wat voor manier dan ook hebben gegeven, en daarmee hebben bijgedragen aan persoonlijke ontwikkelingen die ik in de afgelopen jaren heb doorgemaakt. Hakan: jouw enthousiasme voor het informaticaonderwijs is aanstekelijk. Ik heb veel van je geleerd en ben blij in mijn nieuwe postdoc-positie weer met onderwijs aan de slag te kunnen. Anke: jouw enthousiasme voor TechGirl vind ik fantastisch. Je inspireert me om te gaan voor waar je in gelooft. Ik ben blij dat ik onderdeel van het blogteam ben! Fietsend naar de VU in mijn eerste jaren op als PhD heb ik veel geleerd over feminisme van de podcast Damn Honey. Ik heb daar veel van geleerd, niet alleen over feminisme, maar ook over jezelf zijn. Ook heb ik veel geleerd over genderongelijkheid, maar ook over spreken in het openbaar en het geven van workshops door de organisatie VHTO, waarvoor ik vaak gastlessen heb gegeven. Deze quote geeft precies weer wat mijn motivatie is om altijd te proberen om meer vrouwen in tech te interesseren:

> "I think it's very important to get more women into computing. My slogan is: Computing is too important to be left to men."
> — Karen Spärck Jone

Ik zie dit proefschrift als een prestatie waarvan ik nooit had gedacht dat ik dit zou behalen. Niet omdat ik niet geloofde dat ik het zou kunnen, maar omdat ik niet wist dat het een mogelijkheid was. Ik had (tot ik ging studeren) geen voorbeelden met een wetenschappelijke carrière. Vanaf het moment dat ik begon met studeren, ben ik steeds meer op mijn plek gekomen, mede dankzij de mensen om mij heen. Op een gegeven moment wist ik dat ik een PhD wilde behalen. Dat is me nu gelukt. Ik voel trots, en ook een beetje opluchting. Ik sta voor mijn gevoel nu pas écht aan het begin van mijn wetenschappelijke carrière; waarin ik mijn vleugels ga uitslaan door te gaan werken op een nieuwe universiteit. Na bijna 10 jaar verbonden te zijn geweest aan de VU en te hebben rondgelopen op de campus voelt dit spannend, maar ook verfrissend. Ik kijk ernaar uit om daar een nieuwe weg in te slaan.

> “Altijd, maar m’n ogen op succes
> En altijd, wil ik verder dan het verst
> Maar, ik was er al
> Ik was er al
> Altijd, maar m’n zakken vol met stress
> En altijd, wil ik beter dan m’n best
> Maar, ik was er al
> Ik was er al”
> — Merol, "Ik was er al"

Ik wil afsluiten met de opdracht die ik koos voor dit proefschrift: ‘eet, drink, word gelukkig’; een motto dat ik heb gebaseerd op een titel van een (verder niet bijzonder) boek afkomstig uit de boekenkast van Saskia. Voor mij staan die 4 woorden voor het simpelweg genieten van het leven, zonder dat dat groots en meeslepend hoeft te zijn. Ik ben ongelofelijk dankbaar voor wat ik in de afgelopen jaren heb kunnen bereiken en dat ik dit met jullie kan delen.

*Laura van der Lubbe*

*April 2022 - Hilversum*

# Part I

# Introduction

# Chapter 1

# Introduction

**This chapter contains:**



## 1.1 Motivation

The modern world is full of technological innovation, with a very large potential. Often, technology is only used for the most obvious purposes (such as a phone which is used to communicate in different ways), while its potential is much larger. As technology, and its possible uses, is so diverse, a wide variety of users could benefit from using technologies in many different ways. To understand these possibilities, it is essential to look beyond the technical requirements and specifications. Moreover, it is fruitful to look at the needs of specific groups of people and all processes involved in this. Part of the research of the Social Artificial Intelligence research group of the Vrije Universiteit is aimed at understanding and changing human behaviour, and applying Artificial Intelligence (AI) technology in innovative ways to realize this goal.

In this dissertation, I explore the field of *serious gaming* and *gamification* to evoke behaviour change. Serious games are "digital games created not with the primary purpose of pure entertainment, but with the intention of serious use as in training, education and health care" [11]. Although commercial video games, designed for entertainment purposes, can in some cases also be used as serious games, I exclude those games as they are designed for a general target group and it is difficult to adapt them to suit the needs of a more specific target group. Moreover, non-digital serious games do also exist but are outside of the scope of this dissertation as there are only limited possibilities of including AI technology in non-digital games.

Gamification is different from serious games, as serious games are complete games and gamification is more a description of a process. Gamification can be defined as “the intentional use of game elements for a gameful experience of non-game tasks and contexts” [17]. Gamification can be used to steer users’ behaviour, specifically by using both intrinsic and extrinsic motivation for a lasting effect [5]. For example, a game to change sedentary behaviour can use rewarding game mechanics to create an extrinsic motivation for the user to change their behaviour as they want to earn those rewards. If additionally intrinsic motivation is increased, for example by helping the user to find a sports activity that is enjoyed by the user, the effect on the behaviour of the user can be reinforced.

In 2021 gaming was the biggest earning media sector, potentially boosted by the COVID-19 pandemic and its lockdowns [3]. However, even before the COVID-19 pandemic, games were already increasingly popular [1]. Not only for entertainment purposes but also for an increasing number of domains for serious purposes. Digital game-based learning applications, among which are serious games, can be effective educational tools when applied in a proper and student-centred way [6]. Serious games are applied in many different domains, such as, but not limited to, learning how to program [12], the school curriculum [16], or to improve the health of, for example, older adults [15]. The games are often used as an addition, enrichment, or replacement of existing educational material, due to their specific possibilities and attraction to the audience. Gamification is also used on a very wide scale nowadays, both digitally and non-digitally. From loyalty programs of stores to websites that use game mechanics such as points, levels and rewards to apps encouraging you to move or exercise more with achievements and competitions. Just like with serious games, this dissertation focuses on digital gamification only.

With the work presented in this dissertation, I want to evoke a behaviour change in vulnerable target groups. Societal impact can be made by using techniques such as serious games and gamification to empower vulnerable people. According to the Cambridge Dictionary, vulnerable means “able to be easily physically, emotionally, or mentally hurt, influenced, or attacked” [4]. Following this definition, I denote people as vulnerable in this dissertation when they have some characteristic or are in a certain situation that could potentially negatively influence them (as the definition states). Important is that this vulnerability is caused by something specific, that is not shared with many other people. For example, all human beings profit from eating healthy. However, for some people eating healthy can be specifically challenging and/or beneficial due to some pre-existing health condition. These people can be seen as a vulnerable group. Those vulnerable people can be supported to eat healthy so that they become less vulnerable to the effects of their health condition on their well-being. This support is a form of empowerment. According to the Cambridge dictionary, empowerment is “the process of gaining freedom and power to do what you want or to control what happens to you” [2]. In this context “what happens to you” can be replaced with “what you are vulnerable for”.

Traditionally, empowerment is achieved by human-controlled interventions. Typically, such interventions require participants to move to a physical location and require a time investment of both the instructor(s) as well as the participant(s). However, this can create a burden for possible participants to join. Moreover, it can be hard to reach the right target groups. For example, if you want to reach a very specific subgroup of participants, these might be very divided over a country or region, which makes it hard to reach this subgroup and get them to move to one physical location for an intervention. Digital interventions can overcome such drawbacks. Digital alternatives take away the burden of having to be physically reached and present. Moreover, depending on how the digital intervention is organized, it allows using the material multiple times, which reduces the workload for those organizing the intervention. Of course, for some topics, a digital intervention is less effective or not suitable at all. However, increasing the reach of an intervention can be more valuable compared to the losses when the intervention is changed to a digital version. Furthermore, with the development of new technologies, new possibilities for interventions arise that might have been impossible to move to a digital form before. Altogether, digital interventions such as serious games and gamified applications can increase the accessibility of various types of interventions by both increasing the scalability as well as reducing burdens for participants to join.

### 1.1.1 AI techniques in serious games

For this dissertation, the focus is on how to develop serious games and gamified interventions to empower vulnerable groups, using the possibilities that various AI techniques offer. The goal of using AI techniques in serious games is to optimally assist the users in the game [18], such as creating human-like responses in games to improve believability. Personalization and adaptivity within a serious game can promote motivated usage, user acceptance, and user identification within and outside of the serious games [18]. AI techniques can be used to achieve this. By representing knowledge about users in a player model, adaptations can be made to a game. For example, reasoning about diet constraints and nutritional values to suggest a suitable meal for a user. Together with machine learning, knowledge representation can be used to not only reason about knowledge but also learn from previous interactions. For example, if the system learns about topics that the player struggles with, these topics might get more attention or more points. Another example of adaptation is that machine learning can be used to reason about usage data which can be used to adapt the level of the game so that it remains challenging yet achievable for the user, preventing frustration and dullness at the same time. According to the Flow theory of Csikszentmihalyi [7], balancing skills and challenge keeps players “in the zone” which is between frustration and boredom. Being “in the zone” increases the motivation of players to repeat an activity [7]. Another way to use AI techniques in games is to offer an AI-driven feature that is part of the learning experience. For example, sentiment analysis can be used in dialogues with virtual characters or to analyse the written text of users. Feedback provided by the game can be based on this assessment [20]. In the game *Jobquest*, which is a training game for a job application interview, sentiment analysis is applied to written Curriculum Vitae (CV) to give participants feedback that they can use to optimise their CV [9].

## 1.2 Research questions and dissertation overview

In this dissertation I show how literature and field research can be used in developing serious games or gamified applications for vulnerable target groups. Close collaboration with the target audience is a valuable component in the development of a serious game or gamified application. Studying and developing, and evaluating are often intertwined. This leads to valuable insights that can be used in different research fields. This dissertation also explores what the added value of AI techniques can be, by developing and evaluating new AI algorithms. For example, in the serious game to train resilience against doorstep scams (Chapter 3), voice analysis techniques allow users to also train their tone of voice to increase the assertiveness of their voice. Finally, by incorporating user evaluations, the applications are assessed, and lessons can be learned from that as well. With all these insights, I contribute to research that increases the quality of other gamified applications.

This leads to the following main research question for this dissertation:

> *How can serious games and gamification be used to empower vulnerable target groups?*

This main research question is addressed in four different subquestions in this dissertation, which are discussed in the following parts:

- Part II: *What is the current state of the art for empowering vulnerable people with serious games and gamification?*
- Part III: *How can interactive scenarios be applied in a serious game to make older adults more verbally resilient against doorstep scams?*
- Part IV: *How to design a gamified persuasive diet tracking system as part of a diet intervention for older adults?*
- Part V: *How can gamification be used to develop a self-guided intervention to enhance self-compassion and increase the well-being of young adults?*

The first subquestion is discussed to understand the domain of interest of this dissertation, as well as to establish a common understanding of the important concepts in that domain. Moreover, the current state of the art is studied to understand and identify possible research gaps. This is done with the help of a structured literature review that gives a detailed overview of the current state of the art of the domain of empowering vulnerable people with serious games and gamification. This forms the theoretical basis for subsequent parts. The categorizations made in this part are used to contextualize the other parts in the larger field of serious games and gamification to empower vulnerable people.

The following subquestions each study a different vulnerability and empowerment method. The choice for the domain was often determined by an involvement in a project. The projects together create a wide variety of use cases. I draw general lessons from these use cases. First of all, I draw some lessons about how to design for different age groups that have different technological needs and preferences. Moreover, both a serious game (Part III) and gamified applications (Part IV and Part V) are studied, so my conclusions will address both. All questions address how a digital intervention can be created for something that typically is done in person or manually, however, in each intervention the context and implementation are different.

The first vulnerable group that I study is older adults, both in Part III and Part IV. Older adults are an interesting group, not only because they are vulnerable in different ways (e.g. physically and socially) but also because of the design demands that need to be taken into account to facilitate them to use the technology. It is often thought that older adults are afraid of technology, and do not like or want to use it. However, research by Mitzner et al. has shown that older adults do have a positive attitude towards technology [13]. Moreover, their study showed that ease of use and usefulness are important for predicting technology acceptance. Based on that study, I hypothesize that older adults will be accepting the technologies used to empower them in their vulnerabilities, as it is clear for them how they will profit from them. Other research by Mitzner et al. looked at the technology adoption by older adults for technology specifically designed of them [14]. In their study, it was found that early use of the system, was a strong facilitator for both mid-term and long-term adoption. Those participants who used the system more at the beginning of the trial were more likely to remain using the system to the mid-term and end of the trial. Taking into account the older adults when designing systems is not only important for how they initially evaluate it, but also for their long term adoption. The research from Part III and Part IV contributes to the research towards technology adoption by older adults, more specifically the use of gamified technology, by developing a serious game and a gamified application.

Although older adults are vulnerable in different ways, the focus of this dissertation is on two different types of vulnerabilities, which are addressed by suitable empowering methods. First of all, the safety of older adults can be endangered by doorstep scams. The second subquestion, addressed in Part III, is focused on this topic. Many preventive campaigns focus on more knowledge on preventive methods, however, the focus in the research of Part III is on training their assertive verbal skills to increase their resilience against such scams. Another type of vulnerability is addressed by the third subquestion, discussed in Part IV, namely malnutrition of protein. To empower older adults to increase their protein intake within a diet plan, a supportive gamified application can help them to register their diet and find suitable products within that diet.

In Part V, a different age group is targeted, namely young adults. Those users have other demands, as they are more used to using technology compared to older adults. For this part, the challenge and research contributions have to do with the domain more than with the target group. The vulnerability that I am targeting in this part is the mental well-being of young adults. There are different reasons why young adults do not seek help with increasing their mental health. For example, young adults prefer self-help methods and existing help is often unknown or inaccessible for them [19]. An online, gamified, intervention is a way to make help more accessible for young adults. Mental well-being is a delicate topic, and digital interventions pose challenges when it comes to creating good and sound training material, and personalizing this to the user. The subquestion of Part V focuses both on the design of the intervention. The contributions of this part are in design choices and AI techniques that are used to create an online mental health intervention.

To summarize, Table 1.1 gives an overview of the different parts, their subquestions and the chapters that are used to answer them. Most chapters have been published; the table mentions their references. I did not make any content changes to those publications, except for referencing to 'chapter' instead of 'article/paper' and adding chapter numbers to referenced publications included in this dissertation. For chapters that have not appeared as publications, it mentions the status or publication(s) they are based on.

**Table 1.1:** Overview of scientific contributions of this dissertation

| Part | Subquestion | Chapters & Contributions |
|---|---|---|
| *II. Background* | What is the current state of the art for empowering vulnerable people with serious games and gamification? | Chapter 2: **van der Lubbe, L.M.**, Gerritsen, C., Klein, M.C.A., & Hindriks, K.V. (2021). Empowering vulnerable target groups with serious games and gamification. Entertainment Computing, 38, 1-27. [100402]. |
| *III. Empower older adults to be more verbally resilient* | How can interactive scenarios be applied in a serious game to make older adults more verbally resilient against doorstep scams? | Chapter 3: **van der Lubbe, L.M**, Gerritsen, C., & Bosse, T. (2020). A serious game to improve the verbal resilience against doorstep scams. International Journal of Serious Games, 7(2), 89-119. |
| *IV. Empower older adults for healthy ageing through nutrition* | How to design a gamified persuasive diet tracking system as part of a diet intervention for older adults? | Chapter 4: **van der Lubbe, L.M.**, & Klein, M.C.A. (2019). Designing a system with persuasive communication to improve diet compliance for elderly users. In Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare (pp. 234-241).<br>Chapter 5: **van der Lubbe, L.M.**, & Klein, M.C.A. (2020). Integrating gamification into a system to improve diet compliance for elderly users. In Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good (pp. 150-155).<br>Chapter 6: **van der Lubbe, L.M.**, Klein, M.C.A., Visser, M., Wijnhoven, H.A.H., Reinders, I. Insights on the Effect and Experience of a Diet-Tracking Application for Older Adults in a Diet Trial. Technologies 2022, 10, 31. |

*Continued on next page*

Table 1.1 – *continued from previous page*

| Part | Subquestion | Chapters & Contributions |
|---|---|---|
| *IV. Empower older adults for healthy ageing through nutrition* | How to design a gamified persuasive diet tracking system as part of a diet intervention for older adults? | Chapter 7: Spooren, D., **van der Lubbe, L.M.** Improving the Recommendations of Meals in the PROMISS Application, *Reviewed and Accepted for GoodTechs 2021* |
| *V. Empower young adults to increase their mental well-being* | How can gamification be used to develop a self-guided intervention to enhance self-compassion and increase the well-being of young adults? | Chapter 8: **van der Lubbe, L.M.**, Gerritsen, C., Klein, M. C. A., Hindriks, K. V., & Rodgers, R. F. (2021). Designing a gamified self-compassion training. In A. Veloso, O. Mealha, & L. Costa (Eds.), 22nd International Conference on Intelligent Games and Simulation, GAME-ON 2021 (pp. 79-85). Eurosis.<br>Chapter 9: **van der Lubbe, L.M.**, Gerritsen, C., Klein, M. C. A., Hindriks, K. V., & Rodgers, R. F. (2021). A pilot study of a gamified self-compassion training. In A. Veloso, O. Mealha, & L. Costa (Eds.), 22nd International Conference on Intelligent Games and Simulation, GAME-ON 2021 (pp. 86-93). Eurosis.<br>Chapter 10: **van der Lubbe, L.M.**, Groot, N., Gerritsen, C. (2022). Using Topic Modelling to Personalise a Digital Self-compassion Training. In: Lewy, H., Barkan, R. (eds) Pervasive Computing Technologies for Healthcare. PH 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 431. Springer, Cham.<br>Chapter 11: **van der Lubbe, L. M.**, Gerritsen, C., Klein, M. C. A., Rodgers, R. F. & Hindriks, K.V., Experiences of Users with an Online Self-Guided Mental Health Training Program, *Submitted to Journal of Healthcare Informatics Research* |

### 1.2.1 Research methods

Each project discussed in this dissertation has its specific research method, which depends on the research question, domain and available resources (such as research time or resources to recruit participants) within the project. Overall, similar research methods are used, but the specific approach is different for each project. The domain of serious games and gamification to empower vulnerable groups is a multidisciplinary field, as experts from different research fields collaborate to bring domain and technical knowledge together. In different domains, the conventional research methods might be different. This section serves as an introduction to the general research methods that are used within this dissertation. I also highlight why I think that different research methods are especially important. The specific details for each project are explained in the chapters of the project itself.

Each project consists of two different phases: the development phase and the evaluation phase. The steps within these phases and the relation between the steps and the two phases are explained in this section. Each part of this dissertation addresses these phases in different chapters, which is shown in Table 1.2.

**Table 1.2:** Overview of project phases in the different parts

| | **Development Phase** | **Evaluation Phase** |
|---|---|---|
| Part III | Chapter 3 | Chapter 3 |
| Part IV | Chapter 4, Chapter 5, Chapter 7 | Chapter 4, Chapter 6 |
| Part V | Chapter 8, Chapter 10 | Chapter 9, Chapter 11 |

The **development phase** consists of two intertwined steps: background research and implementation, which together make the design of the application or game. In each of the parts of this dissertation, there is an emphasis on the design of the applications that are built and studied. This is important because it is the foundation of the application, and it gives valuable information to understand the rationale behind the applications. This is important for others, as this enables them to follow the process as closely as possible and to be able to take lessons learned to their research. Without such rationale, it is harder to generalize from the applied case study. In the background part (Part II), it will be described that such rationale is sometimes lacking in papers, while it is an important element when you want to understand how specific choices might have influenced outcomes or to replicate the research following the same rationale.

For the *background research*, different methods are applied. Literature research can be applied to learn more about a domain and the mechanisms that play a role in the studied domain. This can either be done by reading relevant scientific literature, but also by reading other literature such as news articles if the domain asks for it. For example, if the domain is currently in the news, valuable information can be derived from such sources as well. Part III uses such news articles about doorstep scams to study current scam stories. In addition to literature research, field research is valuable when learning more about a domain. Field research can take place in different forms; for this dissertation, it is mainly in the form of interviews with relevant people (such as domain experts or targeted users) in the form of individual interviews or focus groups and studying similar applications. Altogether, these sources form the foundation for the implementation.

To connect the background research with the *implementation* of the application, the requirements for the system are formulated based on the performed background research. This can also be an incremental approach, where you write and validate the requirements in different steps. This is for example used in Part IV: the initial system was built on a specific set of requirements and evaluated in a pilot. Then the system requirements were refined using the input from the pilot and additional requirements from the design of the diet plans. Together with the domain knowledge, these requirements are used to design the application. In general, the domain knowledge and requirements can be used to make choices about which game elements to use. The knowledge can be used in application-specific aspects as well. More technical aspects were designed without (a significant) involvement of domain experts, for example, the actual implementation of the game and AI elements.

When designing systems that will affect human behaviour in some way, it is important to carefully consider ethical aspects that are involved in creating and distributing that system. For Parts III - V, ethical aspects must be considered during the design process. In research in general, an important aspect is ensuring that personal data is gathered only when necessary and stored safely. In addition, as I am dealing with vulnerable groups, it is important to consider that the users will not be harmed by using the system. For example, for the work of Part V this means that the intervention should avoid triggering negative emotions in the users. Ethical considerations will be discussed in the relevant chapters.

Different tools can be used to help designers to take these ethical aspects in mind. The eFRIEND framework [10] is an ethical framework for intelligent environments. Although serious games and gamified applications are not necessarily intelligent environments, they share a lot of characteristics, which makes this framework usable in this case as well. In the discussion at the end of this dissertation, I will reflect on the different aspects of this framework and how I dealt with these aspects for the different developed applications.

The **evaluation phase** also intertwines with the development phase. During the development, evaluation methods can be used to justify or improve the design. After the development, these evaluation methods can be used to assess different aspects of the application, such as the effectiveness or the user-friendliness.

One difference between the projects is the target group. Each target group needs to be approached differently. In the previous section, it was discussed how the design of an application should be suitable for the target group. However, for the evaluation methods, the target group is also a factor to take into account, for example when it comes to recruitment for and communication during the evaluation. For this, you have to adjust to the media that your target group is typically using: older participants might prefer email over chat services like WhatsApp for communication with the researcher, while for younger participants this might be the opposite.

To ensure that I facilitate the target group in the best possible way with the envisioned serious games or gamified applications, focus groups [8] are a very important research method. In focus group research the aim is to gather qualitative data with group discussions or interviews. Often, this is used in the development phase of an application. It can be used to construct requirements for a still to be developed application, but it can also be used to fine-tune a prototype before starting a large scale effectiveness study.

In Part III, focus groups are used not only during the design and development, but also as the main tool in the final evaluation study. Due to the sensitive nature of the topic of the project in this part (doorstep scams), and the difficulty of measuring the effectiveness, it was difficult to design an ethical effectiveness study. Therefore, the main focus is on the experience of participants, for which focus groups combined with questionnaires were suitable research methods as this gives both qualitative data and more in-depth results. Moreover, by combining a focus group with a questionnaire, participants are allowed to share their personal opinions individually as well as clarify important remarks in a group discussion.

In the other parts (Part IV and Part V) an experimental setup is used for the evaluation. In such a setup, different study conditions are compared to gain insights into the effectiveness of an application. However, there is also a strong focus on the experience of participants, which is measured with questionnaires and in different focus group settings. I consider the experience an important factor in serious game/gamification research. Findings on the experience can give context to other found results. For example, if it is found that some results are not as expected, one might find reasons for this in the reported experience. Even if the results are as expected, the additional insights given by the reports on experience can be valuable to know which lessons can be learned from the research. Thus, the evaluations in this dissertation will have a strong emphasis on experience, next to studying the effects of applications if possible.

### 1.2.2 AI techniques to empower vulnerable groups

In the motivation at the beginning of this introduction, I already described different roles that AI techniques can have in a serious game or gamified application: personalization and adaptation, and AI features such as automated feedback with the help of sentiment analysis. AI techniques can be seen as tools to fulfil these roles, to increase the effectiveness and possibilities of the system. The full range of AI techniques can thus be seen as a toolbox that designers of serious games or gamified applications can use. This toolbox is still growing, due to innovations taking place in AI.

In this dissertation, I have applied several tools from this toolbox. All the applications developed in this dissertation are knowledge-based systems. In the section about research methods, I discussed the different ways in which experts and background knowledge are an important foundation for the design of the applications. In Part III, the domain knowledge is translated into scenarios of doorstep scams, together with tailored feedback texts. In Part IV, domain knowledge is used to create rules on how to calculate and visualize nutritional values. In Part V, the metaphors and storyline are based on the domain knowledge acquired, as well as the design for the training and its elements. In all cases, domain experts are involved in the design of these components. Other AI techniques that are explored in this dissertation are voice analysis (Chapter 3), food recommendation algorithms using learning techniques (Chapter 7), sentiment analysis (Chapter 8), and topic modelling (Chapter 10).

During the design process for the different projects, I have gradually added AI features. While sometimes the decision to add a specific feature could be made at the beginning of the project, which was for example the case for the voice analysis in the serious game about doorstep scams (Chapter 3), in other projects possibilities arose later. For example, after the pilot of the self-compassion training (Chapter 9), I wanted to add topic modelling to add another way of personalizing the training. As the toolbox of AI techniques is so limitless, there is no universal way of choosing suitable techniques. This dissertation shows examples of how AI techniques fit different types of serious games and gamified applications.

## 1.3 Contributions

The results of all the projects discussed in this dissertation, combined with the theoretical research in Part II, give a broad view on a specific domain for serious games and gamification, namely empowering vulnerable target groups. The work in this dissertation contributes to understanding the domain of serious games and gamification to empower vulnerable people. The work also contributes to the research on the development of applications within that domain. On top of that, it contributes to understanding how AI techniques can be used to offer (personalized) features that enrich serious games or gamified applications. Finally, for each of the project centred parts, the results that are found in those parts contribute to the research in those specific fields.

## Bibliography


[1] 50 years of gaming history, by revenue stream (1970-2020). `https://www.visualcapitalist.com/wp-content/uploads/2020/11/history-of-gaming-by-revenue-share-full-size.html`, note = Accessed: 2022-01-26

[2] Empowerment - Cambridge Dictionary [online]. `https://dictionary.cambridge.org/dictionary/english/empowerment`

[3] Gaming industry nearly twice as large as reported, at $336b. `https://www.bitkraft.vc/gaming-industry-market-size/`, accessed: 2022-01-26

[4] Vulnerable - Cambridge Dictionary [online]. `https://dictionary.cambridge.org/dictionary/english/vulnerable`

[5] AlMarshedi, A., Wanick, V., Wills, G.B., Ranchhod, A.: Gamification and Behaviour, pp. 19–29. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-45557-0_2

[6] Anastasiadis, T., Lampropoulos, G., Siakas, K.: Digital game-based learning and serious games in education. International Journal of Advances in Scientific Research and Engineering (IJASRE), ISSN:2454-8006 **4**(12), 139–144 (Dec 2018). https://doi.org/10.31695/IJASRE.2018.33016

[7] Csikszentmihalyi, M., Csikzentmihaly, M.: Flow: The psychology of optimal experience, vol. 1990. Harper & Row New York (1990)

[8] Gibbs, A.: Focus groups. Social research update **19**(8), 1–8 (1997)

[9] Gutu, G., Paraschiv, I., Dascălu, M., Cristian, G., Trausan-Matu, S., Lepoivre, O.: Analyzing and providing comprehensive feedback for french cvs with readerbench. Polytechnical

University of Bucharest. Scientific Bulletin. Series C: Electrical Engineering and Computer Science **80**(2), 17–28 (2018)

[10] Jones, S., Hara, S., Augusto, J.C.: efriend: an ethical framework for intelligent environments development. Ethics and information technology **17**(1), 11–25 (2015)

[11] Loh, C.S., Sheng, Y., Ifenthaler, D.: Serious Games Analytics: Theoretical Framework, pp. 3–29. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-05834-4_1

[12] Miljanovic, M.A., Bradbury, J.S.: A review of serious games for programming. In: Göbel, S., Garcia-Agundez, A., Tregel, T., Ma, M., Baalsrud Hauge, J., Oliveira, M., Marsh, T., Caserman, P. (eds.) Serious Games. pp. 204–216. Springer International Publishing, Cham (2018)

[13] Mitzner, T.L., Boron, J.B., Fausset, C.B., Adams, A.E., Charness, N., Czaja, S.J., Dijkstra, K., Fisk, A.D., Rogers, W.A., Sharit, J.: Older adults talk technology: Technology usage and attitudes. Computers in Human Behavior **26**(6), 1710–1721 (2010). https://doi.org/https://doi.org/10.1016/j.chb.2010.06.020, online Interactivity: Role of Technology in Behavior Change

[14] Mitzner, T.L., Savla, J., Boot, W.R., Sharit, J., Charness, N., Czaja, S.J., Rogers, W.A.: Technology Adoption by Older Adults: Findings From the PRISM Trial. The Gerontologist **59**(1), 34–44 (09 2018). https://doi.org/10.1093/geront/gny113

[15] Nguyen, T.T.H., Ishmatova, D., Tapanainen, T., Liukkonen, T.N., Katajapuu, N., Makila, T., Luimula, M.: Impact of serious games on health and well-being of elderly: a systematic review. In: Proceedings of the 50th Hawaii International Conference on System Sciences (2017)

[16] Papanastasiou, G.P., Drigas, A.S., Skianis, C.: Serious games in preschool and primary education: Benefits and impacts on curriculum course syllabus. International Journal of Emerging Technologies in Learning **12**(1) (2017)

[17] Seaborn, K., Fels, D.I.: Gamification in theory and action: A survey. International Journal of Human-Computer Studies **74**, 14–31 (2015). https://doi.org/https://doi.org/10.1016/j.ijhcs.2014.09.006

[18] Streicher, A., Smeddinck, J.D.: Personalized and Adaptive Serious Games, pp. 332–377. Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-46152-6_14

[19] Ustün, T.: The global burden of mental disorders. American journal of public health **89**(9), 1315–1318 (1999)

[20] Westera, W., Prada, R., Mascarenhas, S., Santos, P.A., Dias, J., Guimarães, M., Georgiadis, K., Nyamsuren, E., Bahreini, K., Yumak, Z., et al.: Artificial intelligence moving serious gaming: Presenting reusable game ai components. Education and Information Technologies **25**(1), 351–380 (2020)

# Part II

# Background

## Introduction

To better understand the domain, I start by exploring what is discussed in current research in the domain of empowering vulnerable people with serious games and gamification. Hence, this part answers the subquestion:

> *What is the current state of the art for empowering vulnerable people with serious games and gamification?*

To be able to explain the current state of the art, important concepts need to be clarified. The first sections of the background chapter focus on explaining the concepts from the main research question such as *vulnerability*, *empowerment*, *serious gaming*, and *gamification*. As some of those concepts have multiple, slightly different, definitions in the literature, I motivate the choice of definitions that I used in this dissertation in this first section as well. We propose a classification for serious games and gamified applications from many different domains to understand and clarify how the different domains diverge and where the domain of empowerment for vulnerable people should be positioned in the broad field of serious games and gamification. Moreover, I explain various game mechanics that can be used for serious games and gamification design, which are later used to describe the serious games and gamified applications found in the review.

Next, a structured review is performed to describe the current state of the art. For this, the concepts that have been explained in the first sections of the background chapter are important. The focus is both on the context of the applications as well as on the design and implementation of the applications and the studies. The chapter provides an overview of the state of the art and it offers a structured way of characterizing serious games and gamified applications, specifically for the domain of empowering vulnerable people. Based on this analysis, I identify some research gaps. Finally, recommendations are given to improve research and design in the domain of empowering vulnerable people with serious games and gamification.

To show that it is indeed possible to use the labels proposed in Chapter 2, the introduction of each subsequent part explains how the project described in that part fits the characterization from this background paper, using the terms explained in Figure 2.1 and Figure 2.2.

### This part contains:

# Chapter 2

# Empowering vulnerable target groups with serious games and gamification

This chapter is published as: **van der Lubbe, L.M.**, Gerritsen, C., Klein, M.C.A., Hindriks, K.V.: Empowering vulnerable target groups with serious games and gamification. Entertainment Computing 38, 100402 (2021)

**This chapter contains:**

## Abstract

Serious games and gamification is a popular and growing field, commercially and for academic research. This paper aims to give an overview of a specific domain within the field of serious gaming and gamification; the field of serious games and gamification to empower vulnerable target groups. This overview contributes to a better understanding of this field, by identifying different vulnerable groups and empowerment methods with their own characteristics. From this overview a gap in the existing research can be identified: complex, more indirect, vulnerabilities are not covered in existing research. Moreover, opportunities lie in creating more standardized ways of describing games, enhancing the generalizability of the research. To introduce what distinguishes this specific sub field of serious games and gamification research from other fields, an overview which distinguishes games based on their objective. With the use of a structured literature review, this field is further studied. Next, the identified empowerment methods are studied in more detail, describing the technology, game mechanics and study results found in the literature research. The results of the found studies are often positive, but the generalizability of the results is often limited.


## 2.1 Introduction

Serious games and gamification, the use of games or game elements for serious rather than entertainment purposes, are gaining more and more attention in many fields and are used for various target groups. This increased interest cannot only be seen in the number of applications, but also in the growing research within this area. Academic research studies various aspects of serious games and gamification, often describing specific applications or case studies.

When talking about the types of applications that exist for serious games and gamification, it becomes clear that there are many different domains in which they are used. One way to cluster serious games and gamified applications into domains is based on the goal or application domain. To give an overview of the field of serious games and gamification, this paper starts with a theoretical background section that classifies serious games and gamified applications based on their objectives: the high-level goals of the application (e.g. economic, education or marketing). One of these objectives is of specific interest of this paper: the goal to empower vulnerable target groups with the use of serious games and gamification.

While vulnerability is a concept everybody knows and intuitively understands, it is often hard to explain what the concept actually means [40]. The dictionary provides different meanings, but when studying them closely it is possible to see some common factors. Vulnerable has negative associations. It stems from the Latin word vulnus, which means wound. However, a vulnerable person does not necessarily need to be wounded, he/she has some weak spot that could possibly be a wound when being hurt. So, a person that is denoted as vulnerable has some characteristic or is in a certain situation that could potentially influence the person in a negative way.

Yet it is still hard to pinpoint one definition for vulnerability, as it can be used in different contexts with different appropriate definitions. In order to understand the concept better, a taxonomy for vulnerability can be proposed, explaining three sources and two states of vulnerability [53]. The sources explain where the vulnerability is coming from: inherent, situational, or pathogenic. Inherent sources of vulnerabilities are intrinsic to the human condition; such as hunger or thirst. Situational sources on the other hand, are context specific vulnerabilities; such as vulnerabilities due to natural disasters. Pathogenic vulnerabilities are situational vulnerabilities that are caused by injurious social phenomena. For example: “people with cognitive disabilities, who are occurrently vulnerable due to their care needs, are thereby susceptible to pathogenic forms of vulnerability, such as to sexual abuse by their carers.” [53]. The states explain how pressing the vulnerability is: dispositional (potential) or occurrent (acute). Hoffmaster (2006) explains how vulnerability is in fact harming the autonomy of a vulnerable person. The characteristic or situation causing the vulnerability has taken control from a vulnerable person, the vulnerable person therefore loses its autonomy.

In this research, vulnerability is seen as a condition, that is coming from a personal characteristic (source), that potentially negatively influences the wellbeing or potential of the target group or person. What is affected by the vulnerability is called a vulnerability risk in this paper. To limit the scope of this research, we have chosen to limit this research to personal characteristics (inherent sources), rather than including context specific vulnerabilities (situational sources) as well. Moreover, this paper studies how vulnerable groups can be empowered by serious games or gamified applications. According to the Cambridge dictionary, empowerment is the process of gaining freedom and power to do what you want or to control what happens to you [1]. In the context of this paper, empowerment thus describes the process of mitigating the described loss of autonomy. A serious game or gamified application can be a tool to facilitate this process. Serious games and gamified applications are accessible ways to reach a broad target audience. It has different advantages, among which is the fact that it allows users to independently use it as often as they want. Whereas more traditional empowerment methods, like real-life trainings, might be costly and hard to organize or repeat; developing a serious game or gamified application is a onetime effort. Empowering vulnerable people can have many different benefits, often related to increasing the well-being of the person, but ultimately could also lead to economic benefits, as costs related to the vulnerability are reduced, or the economic opportunities of the person are increased.

This paper studies in more detail how vulnerable groups are empowered by serious games or gamified applications, using a systematic review. In the first part of the review, it is studied which vulnerable target groups are empowered through serious games and gamification. Secondly, the paper identified if there are specific patterns between the empowering method and game aspects such as the type of game, the used game elements or the evaluation method.

The next section discusses the theoretical background to explain the concepts serious gaming and gamification, and introduces a classification of serious games and gamified applications based on their objective. Moreover, this theoretical background explains in more detail what game mechanics, used in serious games and gamified applications, are and which mechanics exist. In Section 2.3, the method of the systematic review is explained. First, in Section 2.4, a review is performed to get insight in the vulnerable target groups and empowerment methods. Secondly, in Section 2.5, the different empowerment methods are reviewed on more technical details and study results. Finally, in Section 2.6, an overall conclusion can be found, giving a summarized overview of the domain of serious gaming and gamification to empower vulnerable target groups.

## 2.2 Theoretical background

### 2.2.1 Defining serious games and gamification

Different definitions of serious games exist; some examples are:

- Serious games "have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement" [2].
- "A game in which education (in its various forms) is the primary goal, rather than entertainment" [58].
- "Digital games created not with the primary purpose of pure entertainment, but with the intention of serious use as in training, education and health care" [50].

The first two definitions do not limit serious games to digital games, as the last definition does. Although it is less common, it is not impossible to have a non-digital serious game. Within this review, the focus will be on digital games. Furthermore, it is also not stated whether the games should be designed for their serious purpose, or that existing games could also be used. However, in all the definitions the emphasis is pointed away from entertainment, which is the primary goal of so-called off-the-shelf-games that are used as serious games. For this review those off-the-shelf-games are therefore excluded. Based on the abovementioned definitions, the following definition that will be used throughout this paper is formulated: *Serious games are digital games, created specifically for an educational purpose and not solely intended for amusement.*

Gamification can be defined as "the use of game design elements in non-game contexts" [20]. Another, slightly different definition describes gamification as "the intentional use of game elements for a gameful experience of non-game tasks and contexts" [76], this definition is used in this review paper. Other variants of these definitions exist as well, some of which focus at using gamification for purposes such as marketing. This review does not focus on such purposes for gamification.

From these two definitions the difference between serious games and gamification becomes clear. Serious games are entire games, whereas gamification is a way to (re)design a non-game task, so it is only adding some elements instead of designing an entire game. Although this distinction might look very clear, it can sometimes be confusing. Within gamification, two different types of gamification can be distinguished: shallow and deep gamification [73]. In shallow gamification the core process is not really changed, but rather enhanced with gamification elements. For the shallow gamification, mainly programming and visual design skills are needed, as there is no need to redesign the main task. This type of applications is clearly different from serious games, as their focus is still very much on the task itself. An example of shallow gamification is adding points and a leaderboard to assignments of a course. For deep gamification on the other hand the core processes are changed for the gamification. In this case game design skills are needed to redesign the activity accompanied by game mechanics, and it can become more similar to a serious game.



### 2.2.2 Classification of serious games and gamified applications

This section proposes how serious games and gamified applications can be classified based on their objectives. One of these objectives is the focus of this review paper, namely empowering vulnerable target groups. This classification can be used to identify whether serious games or gamified applications fit the objective studied in this review.

Serious games and gamification are used for corporate or marketing objectives of commercial institutions. Well known examples of (simplified) gamification for marketing are loyalty programs, in which certain consumer behaviour is motivated by the game mechanics of the program. A customer for example earns points for buying products in a specific store, which eventually leads to rewards such as discounts [15]. Often the goals of these applications of serious games and gamification are of an economic nature (see for example the review of [70]): improving the customer relationship and/or increasing sales. Related to this is the rise of the use of serious games and gamification in tourism (see for example the review of [85]), where serious games and gamification can, for example, be used to increase the satisfaction of visitors with their visit, increasing the chance of them to return.

However, there are also other large domains that have adopted serious games and gamification, without an economical goal but with an educational objective. In the medical domain (see for example the review of [31] or [30]), many different applications exist, for example to train specific situations with health professionals or students. Furthermore, there are many games that are used for rehabilitation purposes [68]. As these games have a specific medical purpose and grounding, we do not include them in the category of empowerment. Another example of such a large domain is education (see for example the review of [55] or [21]), where many different applications are used to motivate and engage students for various different courses or to learn various skills, either physical skills or social skills. In these domains the goal of the serious game or gamification application is not economical, instead it can aim to improve or stimulate education or motivation for learning. However, the target audience is very large and broad, it can be a class following a course, a group of nurses, doctors etcetera. Moreover, although all the target users have something in common (they are students of the same group or have the same job), their personal characteristics, such as previous gaming experience or initial motivation, can differ a lot. Furthermore, many applications have comparable (or the same) subjects, or the application designs for different subjects are comparable. For example, there are many serious games or gamified applications that address a subject such as math or learning a (programming) language. There are also many examples of serious games or gamified applications in the field of education that use game mechanics such as experience points and badges to indicate the students' progress, but for a wide variety of subjects/courses.

There are also applications that do not directly impact humans; for example, applications that try to change the behaviour of users to save energy, improving the climate (see for example the review of [44]). It can be said that these serious games or gamified applications have an environmental objective, as they want to change something in the environment through human behaviour.

Serious games and gamified applications can also be used for research objectives. Data for research purposes is sometimes collected with the use of serious games and gamification. For example the game 'Adventures with Lex', which is used to research the perception from children on the law [8]. However, it can also be used on a larger scale by gamifying the crowdsourcing of certain tasks for research purposes [59]. The main goal of these serious games or gamified applications is to collect data.

The current research focuses on another objective of serious games and gamification; namely to empower vulnerable target groups with serious games or gamification. The target group of such applications is defined by personal characteristic(s), such as age or health condition, rather than a situational characteristic such as a school class or job. Because the subjects of applications with this objective can be very different, this paper aims to study if and which common factors can be found, and explore the area of empowering serious games and gamified applications.

### 2.2.3 Game mechanics

In Section 2.2.1 the following definition of gamification was given: "the intentional use of game elements for a gameful experience of non-game tasks and contexts". In this definition, game elements (or mechanics) are mentioned as discriminating factors between normal tasks and gamified tasks. Game mechanics are not exclusive for gamification, but are used in the broad field of game design. The Mechanics-Dynamics-Aesthetics (MDA) framework for game design [42] aims to guide all the different type of people that are involved in game design (including serious games and gamification) such as designers, developers, and researchers. The components of the framework are:

- Mechanics: the functioning of the components within the game;
- Dynamics: the interaction between the player and the mechanics;
- Aesthetics: the feeling the player gets from the game;

Although both serious games and gamification use these components in their game design, the way it is used differs. As serious games are entire games, all the components of the MDA framework are useful for the game design process. For gamification however, the game mechanics are the main focus.

Within serious games and gamification, game mechanics are used as motivational affordances. Concrete examples of such game mechanics, that are often used for gamification, are: points, leaderboards, achievements/badges, levels, and challenges [32, 87]. In the following paragraphs these game mechanics are explained based on [87].

#### 2.2.3.1 Points

Points are a core component of gamification, either visible to the user, invisible to the user, or any form in between. They can have many different forms, with different purposes, but in essence points are a way to track the behaviour of users and tie reactions to that behaviour. In a game you can use different point systems, either one or more at the same time, to stimulate the motivation of the players. In [87] five possible interpretations of points are discussed:

2

- Experience points (XP): XP is rewarding user behaviour to motivate usage on the long term. By assigning XP to certain activities, the objectives of the designer can be aligned with the player in a long-term way. With XP you can inform your players about what activities are important. XP never maxes out, typically only increases over time (and cannot decrease), and it is not redeemable. Sometimes however, XP can be reset to create goal loops.
- Redeemable points (RP): RP can have many different implementations and names (such as cash or coins). In contrast to XP, RP can fluctuate as players can use their points for rewards. In essence, using RP means that you build an economy, which you have to monitor to ensure it keeps its value to the user. RP systems are useful if you have broad behaviours and/or large groups that you want to motivate. However, it can be challenging to use RP systems for multiple reasons. First of all, there are legal regulations you have to take into account. Moreover, it is important how players perceive the rewards they can get; rewards have to be meaningful and realistic for the users in order for the system to work as motivation.
- Skill points: skill points are assigned to specific activities alongside the core of the game. Users are directed to certain tasks in order to get skill points for specific activities.
- Karma points: Karma points are rare in classic games. The intention of karma points is to stimulate altruism and user reward, the purpose of such points is namely to give them away. They can be used for example as a way to let users reward others for their behaviour in a challenge.
- Reputation points: these points can be used to build trust between different parties that have to work together. It is very complex to build reputation points, as they have to include a wide variety of activities and take into account the different incentives of users. However, in some situations (for example for reviews) it might be necessary to know something about the reputation of a user.

#### 2.2.3.2 Leaderboards

Leaderboards are intuitive for many users to understand, showing them simple comparisons, and putting the earnings of the player in perspective. There are two different types: non-disincentive leaderboards and infinite leaderboards. In a non-disincentive leaderboard the player always sees him- or herself in the middle of the leaderboard, unless (s)he has a high rank. In such cases the leaderboard should show this directly as such a rank might have meaning to the user. If the rank of the player is lower, the emphasis is more on comparing itself with the next-best score for example. Infinite leaderboards are long lists of scores. To make sure that somebody does not fall off the leaderboard different orderings can be used, such as filtering on local or global scores or scores of players with the same level.

Sometimes however caution is needed when using leaderboards, to ensure the privacy of users. It is however not impossible to use sensitive data for leaderboards; it is possible to find ways to do this while ensuring the users privacy.

#### 2.2.3.3 Badges

Badges or achievements can be a way to show accomplishments or progress in a game. Moreover, badges can also serve as a means of social promotion. Badges can be used to replace the level system of a game. Levels are explained in Section 2.2.3.4.

Badges can be implemented in different ways. It might be clear to the user what behaviours result in earning a badge, but this might also be hidden to them. In some cases, it is not even clear what badges exist.

#### 2.2.3.4 Levels

Levels are used to indicate the progress of users. However, nowadays levels are also used in a broader way. In gamification, levels are not used in the traditional way as they were in classic games. When designing levels, you can think of how the difficulty and complexity increases, this could be linear (for example you level-up after earning 100 points), but it could also be exponential or something in between. Creating the levels at a right difficulty and complexity (level balancing) is important to avoid drop-outs, however difficulty can also serve as a way to motivate people as they feel special if they passed a specifically difficult level for example.

Progress bars are a visual way to show users their progress in a level, for example in the form of a percentage of completion.

#### 2.2.3.5 Challenges

Challenges give players a direction for their behaviour. Although it is comparable to badges, it gives more context to the goal that needs to be achieved. Players have something to accomplish within their overall experience. In social games, involving a community of players, cooperative quests or challenges can be created. These challenges are more difficult to build, but also more powerful due to their social component.

#### 2.2.3.6 High-level game mechanics

The abovementioned game mechanics are quite concrete. More high-level mechanics are defined by Schell [75]:

2

- Space: the space of a game is the area in which the player can move around. In this space the player can have different degrees of freedom in its movements. A space can be discrete or continuous, where discrete spaces limit the player to a certain number of possible steps. Furthermore, designers need to think of connections between different spaces.
- Time: timing and length of the game are important mechanics that need to be taken into account for the design process. Using time as a game mechanic can be frustrating, but also motivating. One way in which time can be used is as a constraint, limiting the time that a player can spend on a specific task.
- Objects, attributes, and states: games contain many different objects, all with their own attributes and states. It should be carefully thought out what they are and how they become visible to the user. This also includes what type of knowledge there is in a game, and to what extend the players have this knowledge.
- Actions: this involves the actions that a player can perform and the impact of these actions on the game.
- Rules: rules determine how the game is played. Not only should game designers think of what rules a game has, they should also think about how the rules are communicated and enforced. Furthermore, as described earlier, it is important to make goals in a game achievable and rewarding, or design the right level balancing.
- Skill: games involve different (types of) skills from the players. Which skills and which level of skill is needed can influence the gaming experience? It is also possible to create a gaming experience for users that improves the skills of the players.
- Chance: novelty is a motivating factor in games, which can be created by using chance. However, chance is not always positively experienced by players. For the use of chance, a right balance is important.

## 2.3 Method

This section explains the structure of the review and the concepts used for the review, followed by the process of finding and selecting papers for the review.

### 2.3.1 Review structure and concepts

The goal of this review is to gain insight in the field of serious games and gamification to empower vulnerable target groups. The results of the structured literature study will be presented in the two steps in the next two section. First, Section 2.4 will focus on the *contexts* of the found interventions, rather than the game or research aspects.

The following questions will be answered:

- Which vulnerable target groups are studied?
- How can the target group's risk(s) be characterized?
- In what way do the interventions empower the target group?

Figure 2.1 gives an overview of the terms that are used to describe the papers for those three questions. Each of the above questions is answered with one classification only. In cases in which multiple classifications could fit, the classification that described the most important aspect was used.

With target group we mean the group that the intervention is designed for. In some cases, this can be different from the participants that are used in the evaluation study of the intervention. For example, participants from the same demographic groups as the intended target group, but without a medical condition, to research the user experience without burdening patients. Moreover, the vulnerable target groups are characterized by terms that are mentioned in the paper, rather than doing additional research about their vulnerability. We identify five different target groups, namely people that are vulnerable due to their age, psychical health condition, mental health condition, cultural background or socio-economic status.

The vulnerability risks can be characterized as either health, safety or social risks. Health risks can be related to the current or future health of the targeted user. Safety risks include all types of risks that (could) form a danger to the safety of a person, for example in traffic. Social risks are related to the social wellbeing of a person, for example social exclusion.

With the third question, the focus is on the empowerment methods. For this we do not focus on the technical details, but on the way of empowerment. We identify three different ways of empowerment.

- In order to empower the players, some interventions aim to teach specific skills. These interventions do not aim to teach the player knowledge that could ultimately change its behaviour, to improve its situation, but it teaches a skill that improves the player's situation. Moreover, the intervention does not reward existing behaviour, but rather teaches the user a new skill. We call this empowerment method *training skill*.
- The second category is called *supporting behaviour*. Interventions in this category support the behaviour of the player, both by giving feedback and stimulating specific behaviours. Within this empowerment method, different domains can be identified. On the one hand there is physical activity that can be supported by interventions. On the other hand, there are behaviours, such as substance use or medication usage, that are supported or discouraged by interventions. This type of behaviours will be called lifestyle behaviours in the remainder of this review, and include all other behaviours than physical activity.
- In the third category aim to transfer knowledge to the player. This could ultimately be causing different behaviours of players, but the main goal of such interventions

is to inform the player about a specific topic. This category is called the *transferring knowledge* category.

- When it is not possible to classify an intervention with one of these labels, we will label them with *other*.



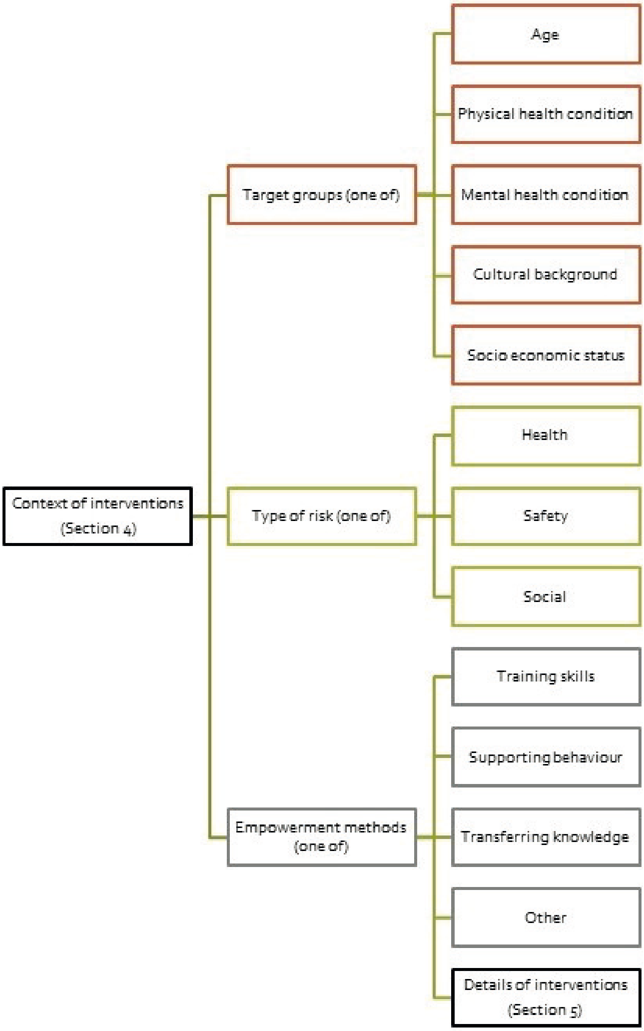



**Figure 2.1:** Overview concepts of taxonomy in Section 2.4

Subsequently, Section 2.5 focuses on the *content* of the interventions and the research. The following aspects are studied in this section:

- Technologies (platforms and hardware) used;
- Game characteristics and game mechanics;
- Types of (intended) evaluation studies and their results.

The technology used for an intervention can be described using the following categories:

- Mobile: applications and games designed for mainly tablets and/or smartphones;
- Desktop: applications designed for computers or laptops;
- Web: web-based applications used in the browser. Sometimes designed for a specific platform;
- Other: this category contains all technology that is used but does not fit under one of the larger categories, such as VR-technology or Kinect.
- Unspecified: when the paper does not mention or is unclear about the technology used

It is possible that an intervention is designed for multiple platforms, these platforms will then all been included in the overview.

The game characteristics and mechanics are described using the following concepts:

- Objectives: some game mechanics are used to steer the behaviour of players within the application or in real life. Such mechanics create a goal for the user to work towards. Examples of these mechanics are: challenges, levels, achievements, and goals. Many of these mechanics are a variation on each other; they all give the player a goal to work towards. Time restrictions or time pressure can also steer the behaviour of the player and the objective for the player is to stick within those time restrictions.
- Rewards: rewarding players is an important component of many games. In some games players receive feedback on their performance, which can be in many different forms such as visually or via audio. However, rewards can also have the form of badges, medals, or prizes. Another way to reward players, is by giving them points. Sometimes these points are used as in-app currency, and sometimes it is only used to show progression or performance. Points are used in many of the games. Another rewarding component can be unlocking something new in the game, for example new content. Finally, rewards are not always positive. Punishment can also be used as a game concept.
- Player data insights: a lot of applications collect data about the player. Interventions sometimes give the player insight in this data, often via a profile page. This sometimes includes a diary function or self-monitoring functionalities.

- Story, scenarios, and avatars: stories, scenarios and avatars can be used in interventions in different ways.
- Notifications: many interventions remind their users of the intervention with emails or notifications outside of the applications. In some cases, these notifications are actual reminders, but in other cases these notifications contain updates about the performance of the user or educational information. As they are delivered outside of the application, it can also serve as a reminder.
- Social features: different social features can be used in interventions. Social features include: teams, user-content generation, interaction with others, social support and competition with others.
- Customization: interventions including customization options in their application.
- Other: some game mechanics cannot be categorized in one of the abovementioned categories, and appear less often. These mechanics are labelled with the category ‘other’.
- Unspecified: when it is unclear what characteristics and mechanics are used, we classify these interventions under unspecified.

Again, interventions can contain different game characteristics. Different categories can be assigned to a single intervention, but when different examples within one category is used, this is only counted once per intervention. It needs to be noted that different papers describe their game or application with another level of detail. It might therefore be that some characteristics and mechanics will not be clearly mentioned in this review, as they were not clearly mentioned in the papers.

To describe the studies and results of the found papers, we study different things. First of all, it is studied whether the reviewed paper includes a study with results, or is a study protocol or design description of an application, without an evaluation. In the case that it includes a study with results, the following aspects are studied:

- Whether the study is a short study (data collected over less than 6 months), or a long study (data collected over at least 6 months).
- We identified different types of studies. An evaluation can be characterized by one of the following types, or multiple if the evaluation consists of multiple studies:
  - Effect oriented studies: these studies offer the intervention to participants for one time and measure the effect of the intervention afterwards. Moreover, other aspects such as usability can be measured as well.
  - Effect oriented studies with multiple interactions: these studies offer the intervention to the participants multiple times, and it measures the effect of the intervention. Again, other aspects such as usability can be measured as well.
  - Experimental studies: these studies distribute their participants over multiple conditions, to compare the results of these different conditions. Experimental studies at least have an experimental condition and a control condition, but can have more conditions.

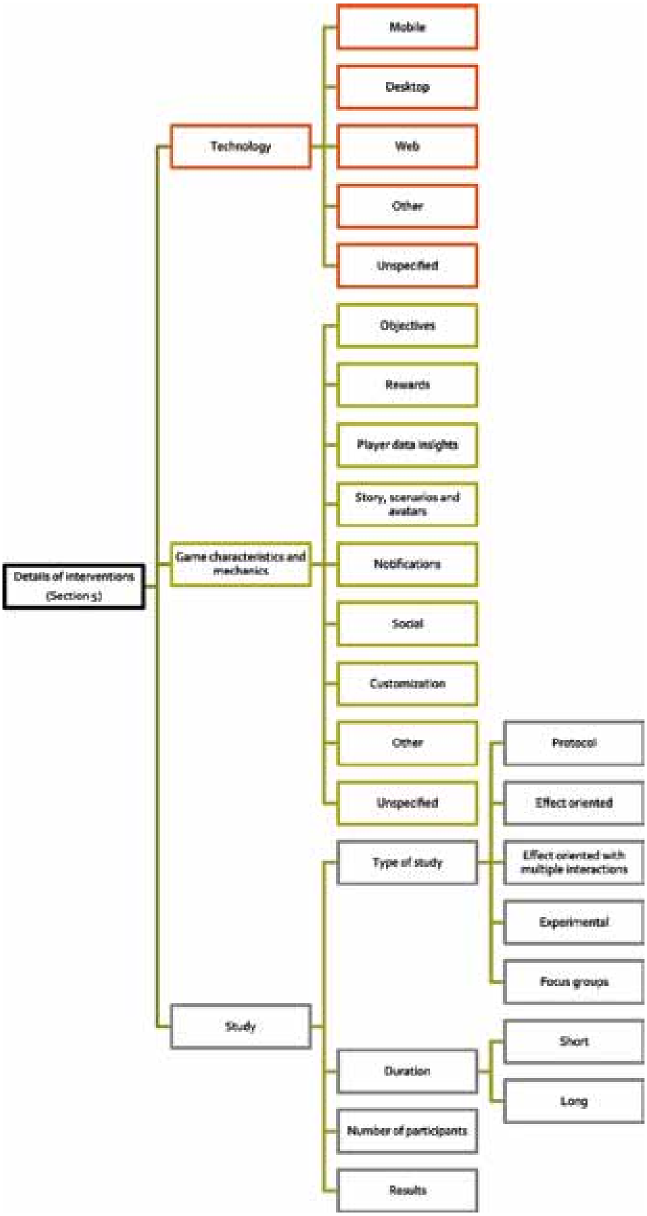



**Figure 2.2:** Overview concepts of taxonomy in Section 2.5

- Focus groups: these studies focus on qualitative data from groups discussions or interviews. In this review we also categorize studies with the main focus on usability testing under focus groups. Focus groups are mainly used to improve a serious game or intervention, or as a pilot study, but not to quantitatively measure the effect of the intervention.



- The number of participants in the study.
- The type of results that are reported.

In some cases, multiple studies are described in one paper. In such cases, the different study types and sizes are mentioned separately.

Section 2.4 describes the different interventions in two different Sections: Section 2.4.1 describes the target groups and their vulnerability, and Section 2.4.2 describes the empowerment methods. In Section 2.5, the different aspects are studied separately for each empowerment method. It is explored whether and, if any, which patterns are found between the findings of Section 2.4, and 2.5. Appendix 2.A summarizes the results from Section 2.4 and Section 2.5 in a table. The only missing concept in this table is *results*, as this is a more descriptive concept that cannot be captured in a table.

### 2.3.2 The search query

As Scopus includes research from many different sources, it is chosen for this review. A comparable search has been performed in the Web of Science database. This resulted in fewer results, but within the most cited papers comparable papers were found. As Web of Science did not result in more sources, it is chosen to use Scopus.

The following query has been executed (on August 2nd, 2019) and resulted in 286 results from the Scopus database:

> PUBYEAR > 2014 AND TITLE-ABS-KEY ((gamif* OR ("serious gam*")) AND (vulnerable OR disabled OR prevent* OR inclusive OR risk OR empower*) AND (evaluat* OR trial OR experiment* OR "user study" OR pilot) AND NOT (rehabilitation OR review)) AND (LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar"))

As we want to focus on recent research, to show the current movements within the field, the query only includes work published after 2014 *(PUBYEAR > 2014)*. We executed the query on the title, abstract and keywords of all references in the Scopus database *(TITLE-ABS-KEY)*.

The first part of the query *(gamif* OR ("serious gam*")* describes the field that is explored within this review: serious gaming and gamification. The asterisk means that words can have multiple endings (e.g. 'gamification' and 'gamify'). To describe the target group, the terms 'vulnerable' and 'disabled' are added to the query *(vulnerable OR disabled)*. Furthermore, four keywords are added to describe the purpose of the applications this review focusses on *(prevent* OR inclusive OR risk OR empower*)*. The 'prevent*' keyword is used because the goal of the application can be to prevent something (for example risky behaviour or an unhealthy situation) from happening. As the application deals with specific groups, inclusive design can be important. Moreover, 'inclusive' also refers to the fact that the applications can contribute to a more inclusive environment for the target group. The vulnerability of the target groups often comes from a certain 'risk'. Finally, the application could target empowerment of the vulnerable groups. As the target group is not always identified as explicitly vulnerable or disabled, it is chosen to combine all the above-mentioned keywords with OR statements.

In order to compare the results of the different researches, or study the effectiveness of different applications, papers are searched in which an evaluation study has been performed. The query includes different keywords that can be used to describe different types of evaluation studies *(evaluat* OR trial OR experiment* OR "user study" OR pilot)*.

Two keywords are used as exclusion criteria *(rehabilitation OR review)*. 'Rehabilitation' is excluded as it is used often in combination with 'disabled', but this combination is not within our field of interest. 'Review' is used to exclude review papers from the query result. Moreover, we limit our results to conference papers *((LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar"))*.

### 2.3.3 Assorting the papers

The titles and abstracts (and if needed the introduction and conclusion/discussion) of all the papers are used to exclude papers that are out of the scope of this research. Figure 2.3 summarizes the process of assorting the papers. After deciding whether a paper should be included or excluded in the review, 120 papers remained. After this duplicates were removed, reducing the number of papers to 103. These papers are used to study the research questions that will be answered in the first part of this chapter (Section 2.4). By answering these questions for each of the included papers, it became clear that although the papers did not meet any exclusion criteria earlier, the papers still did not meet the inclusion criteria. This led to the exclusion of another 44 papers due to different reasons. Some papers were excluded because the target group was not vulnerable due to a personal characteristic, but rather due to a situational characteristic, or the target group was not a specified group. In other cases, the empowered behaviour was not aimed at a vulnerability of the target group, or the target group was not the vulnerable group. Researches describing a game that was not digital, or not primarily designed for the vulnerable target group were also excluded.

After excluding these papers, 59 papers remained that will be used in this review chapter.

## 2.4 Review of context of empowering interventions

This section describes the context of serious games and gamification to empower vulnerable target groups, specifically the target groups, their risks, and the empowerment methods used. Analysing the papers included in the review on the concepts described in Figure 2.1 are used.

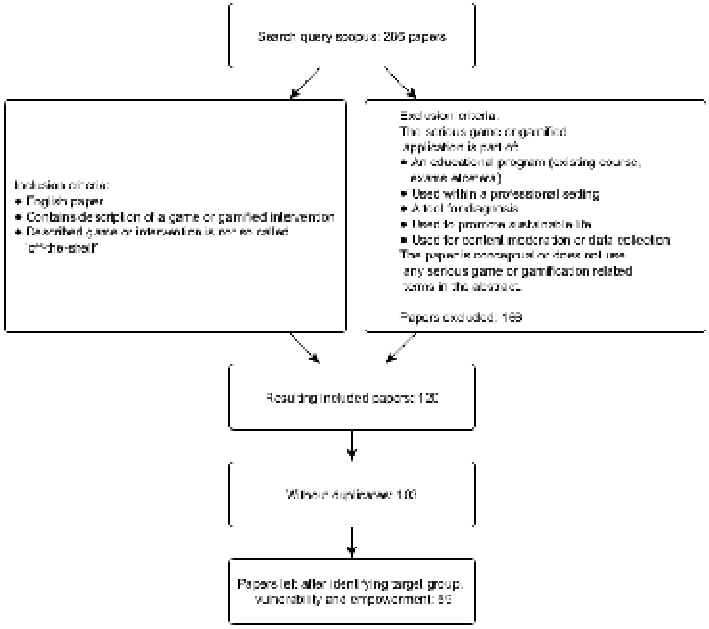



**Figure 2.3:** Review process.

### 2.4.1 Vulnerable target groups and their vulnerability risk(s)

Note that we do not study the underlying relations between vulnerability and the vulnerability risks. In this section we only report the vulnerable groups and vulnerability risks found in the reviewed papers.

The age group was the largest target group. Figure 2.4 shows the distribution of the five different characteristics in the reviewed papers.

Figure 2.5 shows the division of the characterization of vulnerability risks over the papers. From this figure it is clear that 80% of the vulnerabilities are related to health. Although this is a large category, it is found to have different meanings for different target groups. These different meanings will be discussed in later paragraphs in this section.

Figure 2.6 gives an overview of the division of the characteristics of the risks of each vulnerable target group. From this figure it becomes clear that each target group is vulnerable for some health related risks. In the following paragraphs, each target group and its vulnerabilities will be discussed in more detail to give more insights in the different risks for each of the target groups.

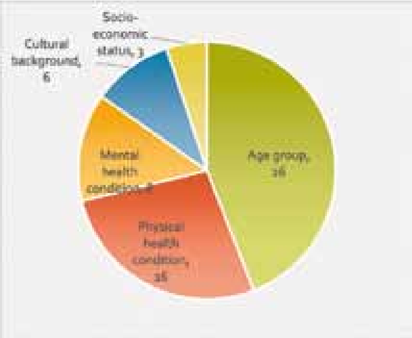



Figure 2.4: Vulnerable target groups.

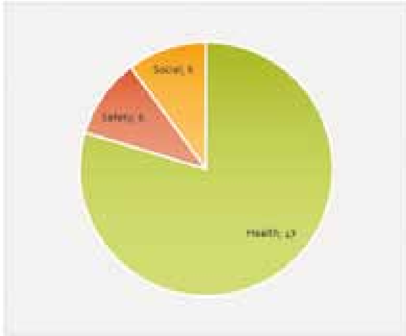



Figure 2.5: Characterization of the vulnerability risks.

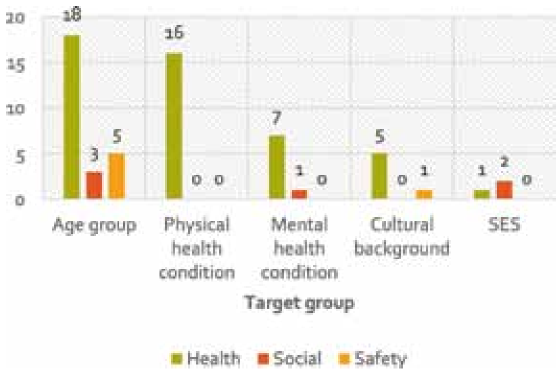



Figure 2.6: Characteristics of risks for the target groups.

#### 2.4.1.1 Age groups

Within the review 26 papers were found in which the target group is vulnerable due to their age [43, 64, 48, 5, 18, 27, 28, 16, 23, 81, 12, 24, 37, 61, 80, 10, 77, 63, 60, 67, 69, 52, 19, 35, 83, 84]. Among these age groups are: children [48, 60, 67, 69], adolescents (with subgroups such as students) [43, 64, 18, 27, 28, 16, 23, 81, 12, 61, 10, 35, 83], adults [24] and elderly [5, 37, 80, 77, 63, 52, 19, 84].

In the papers reviewed, it was found that only the age group was vulnerable for all types of vulnerability risks. This is also shown in Figure 2.6. However, the majority of the papers describe an age related health risk. This concerns health risks that are not coming from or are related to an existing health condition (as these are included in the target groups of health conditions), but the target groups are vulnerable to influencing their health by some behaviour risk that is related to their age.

The following health risks for age groups were found:

- Children and obesity [60]
- Adolescents and substance use [43, 23, 81, 12], depression [64], obesity [28], physical activity [16], eating behaviour [61, 10], HIV/AIDS [35], or bad emotion regulation [83]
- Adults and physical activity [24]
- Elderly and falling [80, 77, 19, 84] or physical activity [63]

A smaller category is the group of papers that describe a safety risk for a vulnerable age group. The following safety risks were found:

- Children and traffic [48] or bicycle safety [69]
- Adolescents and food safety [18]
- Elderly and traffic safety [37] or doorstep scams [52]

Finally, three papers describe social risks of vulnerable age groups: children and their vulnerability for bullying [67], adolescents and their vulnerability for conflicts with peers [27] and elderly with their risk for loneliness [5].

#### 2.4.1.2 Physical health conditions

There are 16 papers in which the target group is defined by a physical health condition [4, 3, 14, 82, 36, 49, 47, 71, 45, 26, 29, 33, 38, 66, 13, 65]. Some papers specify an age group as well, but the physical health condition of these groups is the most distinguishing factor indicating their vulnerability. Physical health conditions that are found are: Rheumatoid arthritis [4], caries [3], Fetal Alcohol Spectrum Disorders (FASD) [14], overweight or obesity [82, 71, 26, 33, 66], HIV-positive [36, 47], people during or after cancer treatment [49, 45], cardiovascular disease [29], and diabetes (both type I and type II) [38, 13, 65].

It is found that in all cases the risk can be characterized as a health risk, as shown in Figure 2.6. In most cases this means that the risk is coming from the physical health condition of the person; for example: children with FASD and disruptive behaviours due to their disorder, or HIV-positive people and the influence of their medicine adherence on their condition. Two papers have a more indirect relation between the condition of the person and health risk they are vulnerable for. In the first paper, people with overweight/obesity are aimed. Their health risk is coming from inactivity; the application focuses on their physical activity [33]. However, the inactivity is not necessarily caused by the overweight/obesity. Another paper in which such a relation is described, is also about physical activity, but among patients with diabetes type II [38]. Again, the inactivity is not caused by the diabetes, but does influence the physical health condition of a patient.

In the age group, the health risks are dispositional; the target group is still healthy, but has a potential risk of harming its health. In this target group, the health risk is more occurrent (while still being a risk and not a certainty) as the current situation of the person is already harming its health.

#### 2.4.1.3 Mental health conditions

Comparable to the target groups in the category physical health conditions, there are eight papers that fit in the category mental health conditions [22, 6, 57, 11, 62, 74, 41, 7]. The following mental health conditions are targeted by these papers: acrophobia [22], Post-Traumatic Stress Disorder (PTSD) [6], addiction: including smokers [57, 11], first episode psychosis [62], dyslexia [74], Obsessive-Compulsive Disorder (OCD) [41], and intellectual disabilities [7].

As shown by Figure 2.6, all papers, but one, describe a health risk. In all these papers, (symptoms of) the mental health condition is the health risk that is targeted by the described intervention. The other paper describes a social risk, namely the risk of social exclusion due to intellectual disabilities [7].

#### 2.4.1.4 Cultural background

For some groups their cultural background contributes to their vulnerability [25, 46, 78, 72, 34, 17]. The following six vulnerable groups in this category were identified: African Americans in the rural Southern United States [25], Chinese adolescents [46], Young people in the Caribbean region [78], Young Africans [72], Adolescents from Sub Saharan Africa [34], and African Caribbean men [17].

Figure 2.6 again shows that all papers but one, describe a health risk. The following health risks have been identified:

- HIV/AIDS for African Americans in the rural Southern United States [25] and young Africans [72]
- Sexual behaviour for Chinese adolescents [46] and adolescents from Sub Saharan Africa [34]
- Prostate cancer for African Caribbean men [17]

The other paper is about a safety risk, namely domestic violence, for young people in the Caribbean region [78].

#### 2.4.1.5 Socio-economic status

For three papers the SES status and/or educational level is the characteristic that makes the group vulnerable [51, 79, 9].

As shown in Figure 2.6, this is the only target group in which the health vulnerability risk is not the biggest vulnerability risk group, but social risks are. The health risk that was found was overweight-related behaviours for adolescents from low SES families [79]. The social risks that are found are parenting related problems for vulnerable parents [51] and the negative effects of unemployment for disengaged and long-term unemployed young people [9].

### 2.4.2 Empowerment

So far, it has become clear which vulnerable groups are targeted and which risks are addressed by the interventions. This section focusses on how the interventions aim to empower their users to overcome or minimalize their vulnerability. The target groups that use the interventions or serious games are from now on also called players or users.


**Figure 2.7:** Division of empowerment methods.

Figure 2.7 shows the distribution of the different categories of empowerment methods that have been identified.

#### 2.4.2.1 Training skills

Within this review, a wide variety of skills that are trained through interventions are found:

- Skill training such as behaviour control, or changing the associations with substances, is used in the domain of substance use [12, 57, 11]. One of these interventions is an exergame in which players have to make physical moves (kicking, hitting etcetera) to counter substance cues [57].
- In the domain of mental health, different skills are trained by the found interventions. Some interventions aim to train skills that can be used to treat or prevent a mental health condition, which is sometimes used in the treatment of the mental health condition [64, 14, 22, 62, 41]. Interventions are also used to train emotional competences or emotion regulation, which is beneficial for the aimed mental health condition [27, 83]. Sometimes games train skills that are less closely linked with the mental health condition. One intervention aims to reduce the loneliness of elderly by motivating and stimulating their motor skills, used to work with computers, by introducing VR to their computer classes [5].
- Next to mental health, also skills related to physical health are trained by interventions. One of the papers researches an intervention for elderly that undergo cancer

treatment, to teach them symptom self-management [49]. A game called Virtual Coach is used for children with diabetes type 1, to teach them about lifestyle adaptations that they can make to have a good life quality [65]. PlayForward: Elm Street Stories is a game for HIV patients to teach them about behaviour and knowledge skills that can reduce the risks of HIV [35].

- To increase the safety of specific target groups, interventions train various skills. For children, the bicycle safety can be increased by training their situational awareness [48], or skills are trained to increase traffic safety in general [69]. Elderly can also be vulnerable in traffic. By training their cognitive functioning, their driving safety can be increased [37].
- With an intervention that includes gamification elements, vulnerable parents' parenting skills can be trained [51].
- IntegraGame is a game for persons with intellectual disabilities, and is used in their vocation training [7].
- For bullying, there is a game that aims to prevent bullying by for example promoting empathy among bystanders [67].
- In order to decrease the chance of elderly to become a victim of a doorstep scam, 'Trucs tegen de Babbeltrucs' aims to train elderly their verbal assertiveness [52].
- Dyslexic children can train their letter recognition with the game Lets find letters [74].

#### 2.4.2.2 Supporting behaviour

It is found that a distinction that can be made within this empowerment method, is about the form of the intervention. Although the technical details of the interventions will not yet be studied and discussed in this section, it has already become clear that there are two types of interventions that support behaviour. First, there are so called exergames, in which the physical behaviour of the user is used as control for the game. Second, there are games that support everyday behaviour. This means that the behaviour of the player is used in a more indirect way in the game. The player does not necessarily be actively playing the game to perform the behaviour (for example walk steps during the day and open the app only once in a while).

Within this review, the following interventions that support behaviour of the users are found:

- Two different interventions support behaviour that is related to substance use. The first game called Alcohol Alert uses the amount that adolescents indicate that they have drank in the game [43]. Another game about alcohol, is a social game concerning peer norms about alcohol usage [23].
- Different games concern overweight or obesity. Their give feedback on behavioural goals of the users or on healthy lifestyle behaviour, sometimes including social pressure as well [82, 26, 79].

- Two interventions support the engagement in care and medication adherence of HIV patients [36, 47]. Both interventions have a slightly different target group.
- When behaviour support can be used to motivate (an increase of) physical activity, for example with step counting or exercises [16, 24, 33, 38]. Another option to support physical activity is using exergames. One research explores whether this could be a way to stimulate physical home training with gamification for elderly [63].
- Eating behaviour is another lifestyle behaviour that can be empowered by behaviour support. In both interventions self-monitoring and feedback plays an important role [61, 66].
- Falling is a risk typical for older people. In order to reduce or assess the risk of falling, exergames are used, these exergames give real-time feedback on movements and thereby support or assess specific behaviours of users related to fall prevention [80, 77, 19, 84].
- Patients with cardiovascular disease have a risk of reoccurrence, and related health risks. To prevent this, MyHeartMate promotes physical activity and other preventive behaviours [29].
- An app, together with a thermal camera, are used for the prevention and management of diabetic foot disease complications [13].
- BUZZING is an intervention to engage young, disengaged and long unemployed, people in their transition into employment. The intervention stimulates experiences that contribute to this [9].

#### 2.4.2.3 Transferring knowledge

Most of the knowledge transferring games are concerning health vulnerabilities, either for existing patients or as preventive measures:

- An informative website with gamification techniques for rheumatoid arthritis patients [4].
- A health education game for children with a high risk for caries [3].
- Games with information about HIV and sexual behaviour, aiming to prevent HIV [25, 72].
- To prevent or reduce obesity and related health risks, interventions educate on nutrition [28, 60] or healthy habits [71].
- An informative game to prevent prostate cancer among African Caribbean men [17].

Next to these points which are closely related to a specific health problem, other interventions were also found in this review:

- Knowledge transferring games about safety. One game is about food safety [18]. Another game is about domestic violence prevention, for young people in the Caribbean region [78].
- Two games are found that educate adolescents on the topic of sexual health [46, 34].
- The game Pure Rush is about substance use [81]. This game has a twofold goal, of which the first goal could also fit the empowerment method of skill training. Namely, to link negative associations with substances and to inform people about drugs.
- One game is about healthy nutrition for university students [10]

#### 2.4.2.4 Other

In some special cases, the empowerment of the target group is not achieved by one the above mentioned methods. The first game within this category is a working memory taxing game that is used for people with PTSD to prevent intrusions [6]. In this intervention, the game is empowering the vulnerable target group in a way that it is a tool for them; the game is used in a moment to directly reduce their vulnerability.

The second game in this category is the INTERACCT serious game [45]. This is a communication tool with gamification that is used for children after their cancer treatment, used in the communication with care takers. It is used within the care after their cancer treatment, while for the children it can also be seen as an entertaining tool.

### 2.4.3 Conclusions about target groups, vulnerability risks, and empowerment

In Figure 2.8 it is shown how the empowerment methods are used for the different vulnerability risks. This figure shows that only for health risks all empowerment methods have been used in the literature reviewed. Most popular within this category is the support of behaviour, followed by transferring knowledge. For safety risks mostly training skills and transferring knowledge is used. For social risks this is comparable, with the only difference that instead of transferring knowledge, supporting behaviour is used.

In Figure 2.9, an overview of the target groups, vulnerability risk categories and empowerment methods is shown. This figure shows that age group is the most diverse category, having both the most diverse types of risks as well as using the most diverse set of empowerment methods. What stands out is that supporting behaviour and transferring knowledge are almost exclusively used for health risks, of mainly age groups but also for the physical health condition and the cultural background groups.

The target group physical health condition is only empowered in their health risks. In the mental health condition target group, there is one intervention aimed at a social risk (social exclusion), instead of all risks being a health risk. For physical health conditions, one would also expect the presence of social risks, however those risks are not found

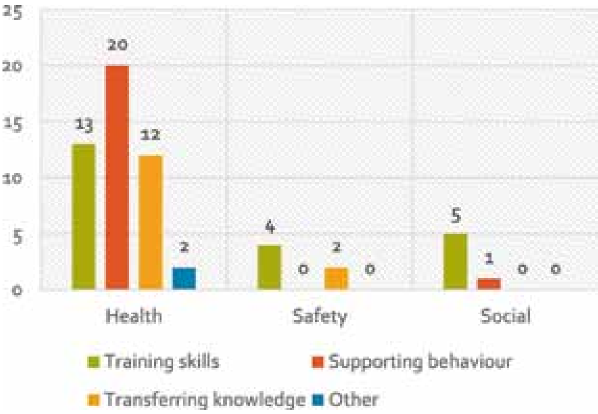



**Figure 2.8:** Empowerment methods for different vulnerability risks.

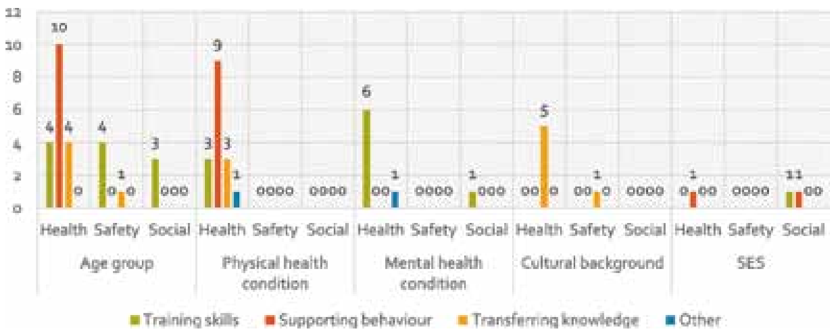



**Figure 2.9:** Empowerment methods for different risks for vulnerable target groups.

within this review. Risks that are targeted by the interventions are closely related to the current condition of the person, but less closely related risks are not addressed in the found literature. In the introduction the term pathogenic vulnerability is explained. It seems that these types of vulnerabilities are not often addressed by the described applications. One example of such a vulnerability that is addressed is the domestic violence in the Caribbean [78]. However, other larger themes of pathogenic vulnerabilities are not found.

## 2.5 Review of technologies, game characteristics, and study results

So far, we studied the papers in the scope of our review on a high level to give an overview of the serious games to empower vulnerable target groups. However, this only says something about the contexts of the intervention, but not about the specifics of the serious games and applications. This section will investigate the papers on a detailed level, to give an overview of the following specifics of the research reported on:

- Technologies (platforms and hardware);
- Game characteristics and game mechanics;
- Types of (intended) evaluation studies and their results.

Figure 2.2 shows which concepts are used to describe these three elements. In the remainder of this section, the three points are analysed for each empowerment method. Finally, for each point the general conclusions are drawn.

Note: one of the included papers could only be studied for the first part of this research, as only the abstract and introduction could be accessed by us [46]. This paper falls in the categories: cultural background, health risk, transferring knowledge. In this second part of the review, this paper is excluded.

Moreover, note the following for the sections about transferring knowledge: The game used in the research of [24] is previously described in a pilot study [54]. For the remainder of this section, the reference of the paper found in this review will be used. However, some information comes from the earlier paper. Similarly, the game described in [38] refers to a previous paper about the game [39], which is also used for the review of the game mechanics.

### 2.5.1 Technologies

#### 2.5.1.1 Technologies for training skills

Some games can be played on a mobile device (smartphones or tablets) or a device with a touch screen [48, 37, 52, 22, 11, 41]. Sometimes additional hardware is needed, such as VR goggles, for a smartphone application [22]. Other games are played on a computer or laptop [27, 12, 69, 14], are VR computer games [5], or other categories such as exergames played with the Kinect technology [57]. One of the videogames within this empowerment method is used in the classroom to empower students to overcome social conflicts [27]. Another game that is played on a school computer, is an online game for depression prevention that is used together with a paper notebook [64]. For one game, the physical device was not specified, but the target platform has been mentioned, in this case web browsers [62].

Games can also be suitable for different platforms. A game to prevent bullying is designed for both desktop as well as tablet devices, and can be used both in a private session or in a session with a therapist [67]. Another gamified intervention for parenting skills has a responsive design so that it can be played on both mobile and desktop devices [51]. There is one VR game that is tested on three different devices: computer, smartphone, and rgb-d camera [83]. However, there are also some papers that do not specify their target platform [35, 49, 65, 74, 7].

Figure 2.10 shows how the different technologies are used for the different target groups and their vulnerability risks. All technologies are quite scattered over the different groups.

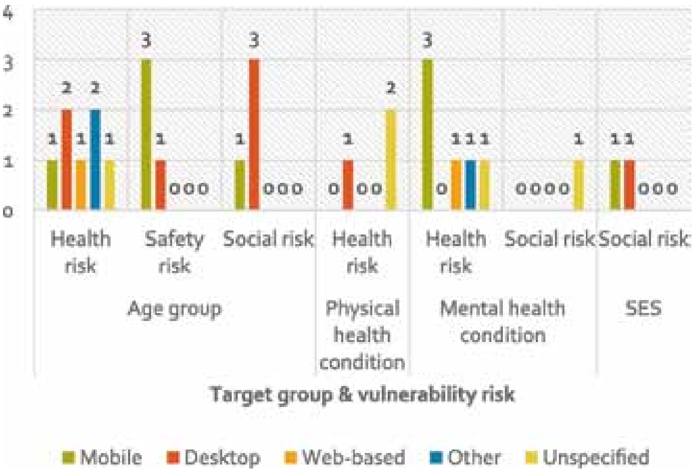



**Figure 2.10:** Technologies used for training skills in different target groups.



#### 2.5.1.2 Technologies for supporting behaviour

Most interventions are mobile applications [23, 24, 84, 82, 36, 26, 29, 38, 66, 13, 79, 9] sometimes using the sensors of a smartphone or wearable [29, 38], non-invasive wearable sensor [84] or an additional thermal camera [13]. One of these interventions also includes an online platform [9]. Two interventions can be played via a website, but are both accompanied by some additional material: a workbook [16] or a wearable device to track the steps and sleep of the user [33]. Another game accessible via a website, is played on a computer [43], while another game mentions different target devices as an online game [61].

As discussed in Section 2.4.2.2, exergames are also used to support behaviour of players. For these exergames slightly different technology is used, such as Kinect technology [80, 63], wearable devices [77], or Wii Balance board [19]. Moreover, these games are often played on a larger screen or television screen. These technologies belong to the category ‘other’ as they are only used in a limit number of applications and do not belong to the other categories.

Another paper [47], does not specify which target device or platform is used.

Figure 2.11 shows that mobile apps are the most popular platform, and are used for each target group. Mostly they are used for health risks in the physical health condition group. Exergame technology, labelled as category ‘other’, is only used for age groups with a health risk, more precisely for elderly and mostly for fall prevention.

#### 2.5.1.3 Technologies for transferring knowledge

Most applications that aim to transfer knowledge to their players, are web-based [18, 28, 81, 10, 4, 25]. Most of these games do not specify their target device. However, one of these games is meant to be played on a computer [28]. One of the games is not exclusively for web browsers, but it can also be found as an installable app [81].

Other games are either meant for mobile devices, smartphone [72] or tablet [3], or desktop/computer [60, 78, 34]. One of these computer games is only played at school [34].

One intervention consists of multiple components, including apps and commercial material such as the game Wii Fit Plus [71]. Finally, one intervention does not mention its target platform [17].

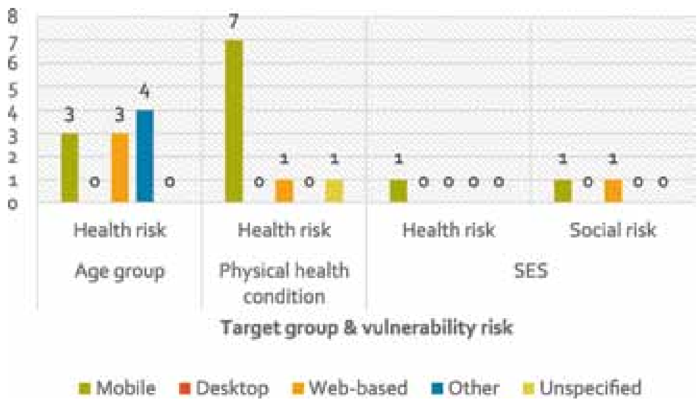



**Figure 2.11:** Technologies used for behaviour support in different target groups.

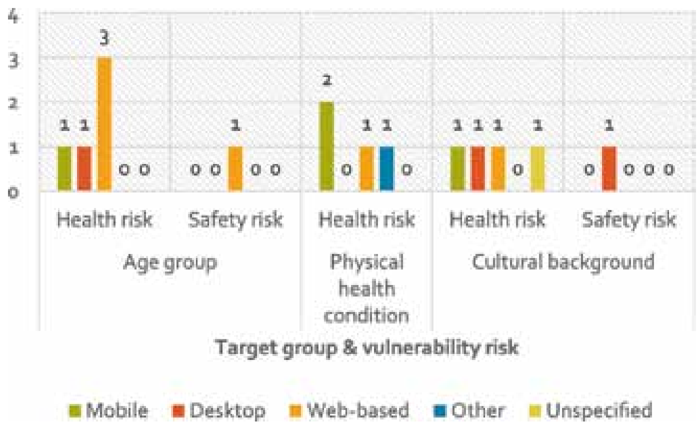



**Figure 2.12:** Technologies used for transferring knowledge in different target groups.

Figure 2.12 shows that web-based applications are mostly used for health risks, but also for the safety risk of the age group. The other technologies are scattered over the different target groups and risks.

#### 2.5.1.4 Technologies for other interventions

The first application from the other category is a communication tool for children after cancer treatment [45]. This is a web platform that uses data gathered with a Kinect or Android device.

The second game in this category, is a working memory taxing game, used for people with PTSD to reduce intrusions [6]. The platform or target device of this game is not defined.



#### 2.5.1.5 Conclusions about technologies

In total, 66 types of technologies[1] are used by the 58 papers that are included in this part of the review. Figure 2.13 shows that the biggest category is mobile technology, making 38% of all technologies used. For both the empowerment method of training skills as well as for supporting behaviour, mobile technology is the preferred platform.

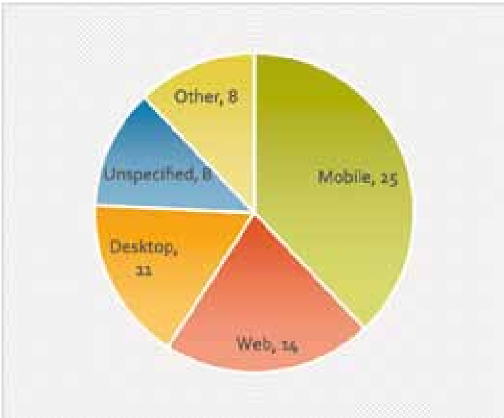



**Figure 2.13:** Technologies used for all empowerment methods.

Training skills moreover uses the desktop platform often, supporting behaviour on the other hand does not use the desktop platform at all. Exergames, often using 'other' technology, are only found for the supporting behaviour category.

### 2.5.2 Game characteristics and game mechanics

#### 2.5.2.1 Game characteristics and game mechanics for training skills

**Objectives** Steering the behaviour of users can be done using challenges, missions, targets, tasks, goals, or levels [64, 12, 37, 67, 35, 14, 22, 11, 74, 7, 51], sometimes with increasing difficulties. Another way to steer the behaviour of a player in the application is

[1]When more than one platform within the same category is used, this is counted as one.

by using time pressure, which can be in the form of limited time (for specific levels) or a timer [69, 83, 11, 74].

**Rewards** In some games this feedback is provided in a visual way [48, 12, 37], as advice [22] or via audio [74]. Other papers mention giving feedback, but do not specify the type of feedback [67, 83, 14, 51]. One game gives both visual feedback as well as textual feedback, on different aspects of the users choices and behaviour in the game [52]. Another type of reward can be unlocking new modules [22]. However, points [12, 37, 67, 69, 52, 65, 57, 62, 74] are also a way to give feedback on performance or behaviour. Points can have different names and values. Furthermore, points can be represented with numbers, but also in a visual way, for example in the form of a glucose bar [65] in an game about diabetes. Other forms of rewarding feedback are badges and progress bars [69, 51].

**Player data insight** Giving players insight in their data and performance can be another form of feedback [69]. Another game includes a profile showing the players information about themselves [22].

**Stories, scenarios, and avatars** Some games use a story or scenario component in their game, for example to give players insight in the consequences of their actions [27, 12, 67, 52, 35, 49, 22, 62]. Sometimes these stories consist of different levels that can be unlocked consecutively [67, 22]. In another scenario game, players can choose between different profiles (bystander or victim), resulting in different games. Moreover, different characters can be chosen [67].

While in most of these scenario games, it is about making choices and progressing through the scenario which ultimately teaches the player a new skill, one game also has an additional component. In this game, players also have to use their voice to influence the progress of the scenario. One of the skills that is taught by this intervention is an assertive way of using the voice, which is trained with this component of the game [52].

One game uses stories in a different way. In OnTrack > The Game, stories of hope and recovery, told by people who have already experienced first order psychosis (the target vulnerability of this game) are included [62]. Next to this, the player can customize their character's appearance and explore an imaginary world. In this world they can have experiences that provides them with skills that help them to deal with their real world problems.

Another way to give a player insight in their behaviour, and learn them about behaviour that suits their physical health condition (in this case diabetes type I), is by using a virtual avatar that the player needs to take care of [65]. In this game not only a virtual avatar is used to take care of, but another virtual avatar has the role of coach. Other games that use an avatar, or user character, is a game about traffic safety [69], an anti-smoking game [11], and an intervention for vulnerable parents [51].

**Notifications** Notifications in the form of emails to stimulate the adherence of participants is used in one intervention [22]. Reminders for an intervention are also used in [51].



**Social** Social (media) features are used in an intervention for vulnerable parents [51].

**Customization** As mentioned before, the character of the player in OnTrack > The Game can be customized [62].

**Other game mechanics** Two games make use of mini-games [35, 41]. Moreover, one paper describes an intervention consisting of two different games for emotion regulation strategies for adolescents [83]: one game in which you have to pop as many balloons as possible, within a time limit. The second game is a game which teaches the player breathing strategies: by breathing in a specific pattern, the player controls a feather through the game.

An intervention for vulnerable parents also includes professional moderation in the network of the intervention [51].

**Unspecified game mechanics** For one paper the game characteristics and mechanics were unclear. This paper is about preventing loneliness among elderly by using VR in their computer classes [5]. The paper describes various games, all with their own mechanics, that are used in this study, but the games were not designed within this study, but rather reused from others.

**Summary** The most used game mechanic categories for training skills are[2]: objectives (including challenges, targets, tasks, levels and time pressure), rewards (including points, feedback and badges), and stories, scenarios and avatars. In Figure 2.14, it is shown how often these categories are used. Rewards is by far the most popular category. All three popular categories are used in more than half of the interventions.

Figure 2.15 shows for the different target groups, in how many of the games the most popular game mechanics are used.

#### 2.5.2.2 Game characteristics and game mechanics for supporting behaviour

**Objectives** Behaviour support applications are built around mechanics to motivate behaviour of the player in real life through game mechanics. Mechanics such as goals (setting), tasks, and planning [61, 36, 29, 38, 79, 9] play a role in this. Chance [23] and challenges (performed in the real world) or achievements [43, 24, 82, 36, 47, 29] are other game mechanics that can be used for this. One game challenges the user by showing its streak of medication usage [47]. Levels can also be used as a way to stimulate repetitive usage of the app, and it is used in different applications [16, 19, 36, 29, 33, 66]. Also a progress bar can be used to motivate users [13]. Finally, one of the exergames uses time in some of its mini-games to ensure that the exercises are performed in the right way (fast

[2] >5 times, excluding the category 'other'.


**Figure 2.14:** Popular game characteristics for training skills.

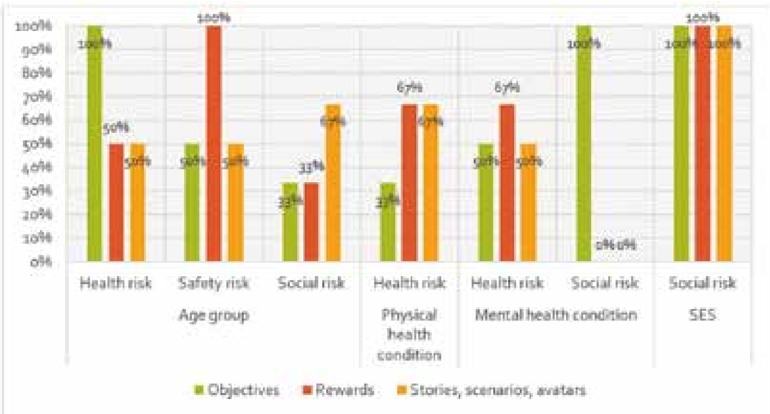



**Figure 2.15:** Popular game mechanics used for training skills for different target groups.

or slow) [77]. Time can also be used in another way, when high scores are recorded to stimulate players to improve their own performance [19].

**Rewards** Games give players feedback on their performance [43, 23, 61, 77, 63, 19, 26, 29, 38, 66, 79]. Feedback can also be given to the player by giving them rewards, which is done in different forms [16, 24, 36, 47, 26, 29, 13, 79], for example with badges or trophies. One game gives tips written by comedians as reward to the player [24]. Unlocking levels, features or mini-games in the game can also be a way of rewarding players [19, 36, 47, 66, 13]. However, negative rewards, in the form of punishment, are also used [26].

Another form of rewards is points [23, 82, 47, 26, 33, 66] or in-app currency points [36, 29, 38] can be used to reward the player's performance as well.

In one game players need to clean and improve a garden [38]. It was mentioned that the choice for a garden was made as previous research pointed out that this can have relaxing rewards on users.

**Player data insights** The component of giving players the option to track their behaviour and give insight in their data, for example via a user profile, is also used in different behaviour support applications [24, 61, 82, 36, 47, 29, 38, 66, 13, 9]. In one case, the app uses an additional thermal camera to collect real world data [13].

**Story, scenario, and avatars** There are three papers that use the mechanic of a story to achieve its goal, namely to motivate the players (HIV patients) to stick to their treatment and give them support to this behaviour [36], or a storyline in a game about healthy eating [66]. Finally, in an intervention about binge drinking, players need to find out what happened to their character [43].

One of the exergames for fall prevention describes the user-centred design process that is used to explore what a game for the purpose of physical activity among elderly should look like and include [63]. Prototypes are created for evaluating different aspects, it is described that part of it will be an avatar that gives motivational feedback and tips.

Another game that uses an avatar is a game in which the player needs to take care of a virtual avatar [29]. As the target group for this app is people with cardiovascular disease, this avatar is in the form of a heart.

**Notifications** Many applications use notifications to enhance the usage of the application [43, 36, 29, 38, 13, 79]. Some applications send the player weekly updates about its performance, which is also a way to remind them of the application [24, 33]. Another application sends their users daily tips and messages [66], while another reminds its users to take their medication and gives them a daily newspaper [47].

**Social features** Many of the games of this empowerment method incorporated a social component. However, the games used this in different ways. It can be that players are visible to others [23], that there is a leaderboard [29], or that players can interact with other players [47] or for user-content generation [23]. Various games also include competition [16, 23, 26, 79].

However, more often it includes social (media) support [36, 26, 33, 38, 79, 9] or teams [16, 24, 77, 26, 33]. An example of a game that uses teams, makes these team members report to each other on a daily basis [26]. Components like peer support, accountability and team discussion, are used within this intervention.

Another intervention uses both the social component of using groups, and individual components [16]. There are competitions for teams of players, and individual users can play different levels. Moreover, teams can earn trophies and both schools, teams, and individual users can earn different types of real-world prizes for their performance. Comparable, in the intervention described in [24] players can earn rewards for their individual performance, as well as their team performance. Players can also exchange gifts with their team mates. One of the theories that is used as foundation for this application is the fun theory [86]. This is represented in for example the tips that players receive. These tips are written by comedians. Comedians also picked the names for the awards that players and teams can unlock.

One intervention does not only provide online support, but also organizes activities outside of the application [9].

**Customization** Three papers describe using a personalisation approach. In the first application the player is provided with games and activities that have an educational value in the domain of weight loss [82]. The second paper adjusts the level of its exercises to the level of the player [38]. Personalisation is also used by a game in which the players can customize their avatar's appearance [47].

**Other game mechanics** One of the exergames for fall prevention consists of three mini-games for two teams of maximum three players. Each game has its own rules and mechanics. The overall theme of the exergame is Antarctic, there are many penguin avatars used. To control the games, the players need to perform different exercises. The wearable device that is needed for this exergame, can be used in different ways, depending on the game and exercise [77]. Mini-games are also used in another game [29].

**Unspecified** Some interventions do not specify their game mechanics. The intervention includes balance tests and exercises, based on existing tests and exercises [84].

One of the exergames consists of different exercises, that can be performed under the supervision of a therapist [80]. However, the exercises are not further described in the paper.

**Summary** The most frequently used[3] types of mechanics for supporting behaviour are objectives (such as goals, levels, challenges), rewards (such as feedback, badges, points), social elements, and notifications and player data insights.

Figure 2.16 shows how many games of the empowerment method of supporting behaviour use the most used game mechanics. This shows that rewards and objectives are by far the most popular mechanic in this category. Social aspects are used in more than half of the games; notifications and player data insights are used in slightly less than half of the games.

Figure 2.17 shows the percentages of games from each of the vulnerable groups that uses the most popular game mechanics. This figure shows that especially the physical health condition group with a health risk uses all the popular mechanics very frequently, the SES groups as well, but these only contain one paper.

One of the papers of this empowerment method is a good example of how a paper can clearly explain the game mechanics and the rationale behind the choices for these mechanics [16]. The paper includes a table that explain the intervention components and connected behaviour change techniques. By doing this is a table, it is easy to get a clear and complete overview of all components and rationale behind these.

[3] $\geq$10 times.

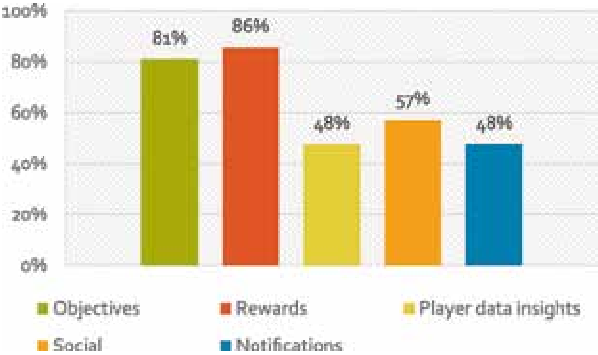



**Figure 2.16:** Popular game characteristics for supporting behaviour.

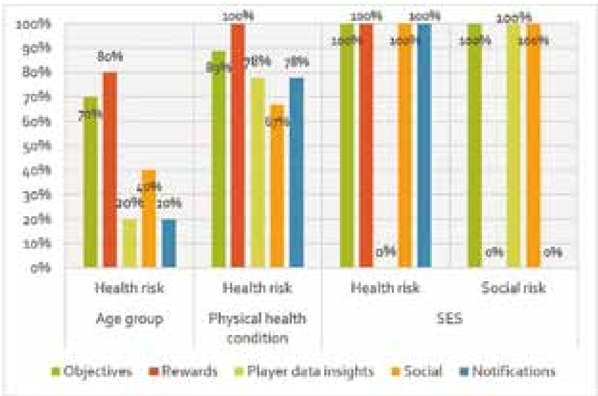



**Figure 2.17:** Popular game mechanics for supporting behaviour used for different target groups.

#### 2.5.2.3 Game characteristics and game mechanics for transferring knowledge

For one intervention in this empowerment method [28], the game is described in a previous paper [56]. In addition to the example mentioned at the end of the previous section, this is an example of a game description that is very clearly and schematically presented to the reader in a table. In the remainder of this section we will use the reference from the paper found in this review is used (reference [28]), while information might come from the previous paper.

**Objectives** Interventions with the purpose to transfer knowledge use, just like all other methods so far, game mechanics to steer the behaviour of the user through the intervention, such as missions, levels, time restrictions [18, 28, 71, 78, 17].

**Rewards** Rewards and feedback are used in different ways to reward users during the game [18, 28, 81, 10, 4, 71, 25, 34]. Examples of rewards use are points, gifts, medals, badges, and prizes.

**Player data insights** One intervention is registering physical activity and eating behaviour [71]. Another game gives personalized advice based on the game [17].

**Stories, scenarios, and avatars** Different games within the category of knowledge transferring games use stories or scenarios. Often, players influence the story with the choices they make. In a point-and-click game about domestic violence, users can interact with different characters and play different levels [78]. Another game, about prostate cancer, also uses characters and levels [17]. A 'choose-your-own-adventure' game, about HIV, is designed in such a way that the player can replay it, to see how different choices influence the story [72]. Another game with story elements is about sexual health [34]. Stories and questions are used to transfer knowledge to the player in this game. In a game about food safety, players have to solve a mystery and play different mini-games to do so [18]. Avatars are also used in a game for drugs education [81].

'Space adventures' is a game made for children to learn about healthy eating to reduce the risk of obesity and related health risks [60]. The game mechanics of this game are not explained in much detail. It is mentioned that they use a narrative component, and on the pictures of the game it is shown that the game also has characteristics of a platformer. Players have to make choices about food in the game.

**Notifications** One intervention sends regular emails to inform players about their performance [10].

**Social features** Social features are used by different informative platforms. Competition is an important aspect used in different games in the form of, for example, leaderboards [10, 4, 34].

**Customization** Avatar appearance or gender can be customized in two different interventions [81, 72].

**Other game mechanics** It was already mentioned that questions or choices were used in stories, but there are also interventions that use it in the form of quizzes [10, 4, 34]. One intervention includes different games and apps in order to prevent childhood obesity [71]. Mini-games are also used in [18, 72].

**Unspecified** One paper is not discussed in this section [3], this game is an improved version of a previously described game. However, it was not possible to find the game mechanics of the game.

**Summary** The most popular categories[4] of game mechanics for transferring knowledge are objectives (such as levels, missions and time restrictions), rewards and feedback (including points), and a story element.

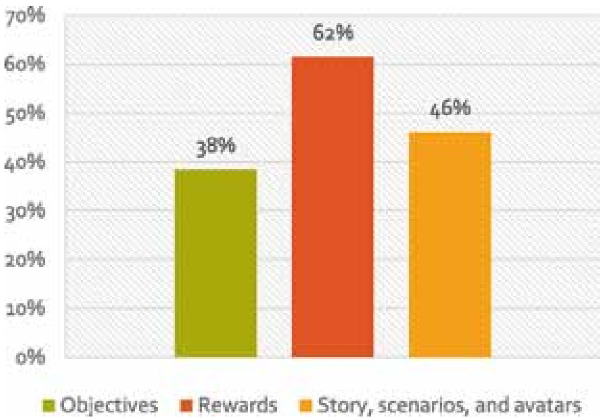



**Figure 2.18:** Popular game characteristics for transferring knowledge.

Figure 2.18 shows that only rewards are used in more than half of the interventions, the other popular mechanics are used in less than half of the interventions.

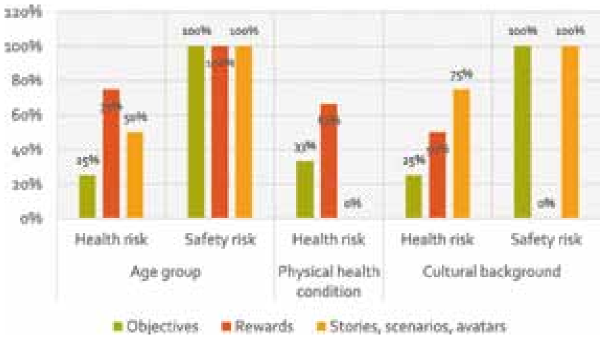



**Figure 2.19:** Popular game mechanics for transferring knowledge used for different target groups.

Figure 2.19 shows the percentages of games within each vulnerable target group using the popular mechanics. This shows that the games with a safety risk for an age group include all three mechanics. Rewards is a very popular element for age groups with a health risk or people with a physical health condition and a health risk. However, for vulnerable people due to their cultural background, this mechanics is less popular.

[4] >5 times, excluding category 'other'.

#### 2.5.2.4 Game characteristics and game mechanics for other interventions

The first application, for children after their cancer treatment, explains how it serves different purposes [45]. On the one hand it is a form of entertainment for the children but this is not the only purpose; the platform can also be used to facilitate the communication between the child patient and the professionals. The game uses challenges (objectives), games (other), social elements and competitions (objectives) to achieve its goals.

In the second research in this category two different experiments were performed with two games to reduce intrusions after a traumatic event [6]. The first game is a 'collect and avoid' game with different levels with time limits and stars to collect during the game. The second game is a Tetris like game, but instead of making lines the player has to make squares.

#### 2.5.2.5 Conclusions about game characteristics and mechanics

In total, 168 different implementations of categories of game mechanics are used in the 58 interventions included in this section. From Figure 2.20 it is clear that rewards (24%) and objectives (23%) are the most often used mechanics. These categories are also the most diverse categories, as many different types of game mechanics are mentioned in these categories.

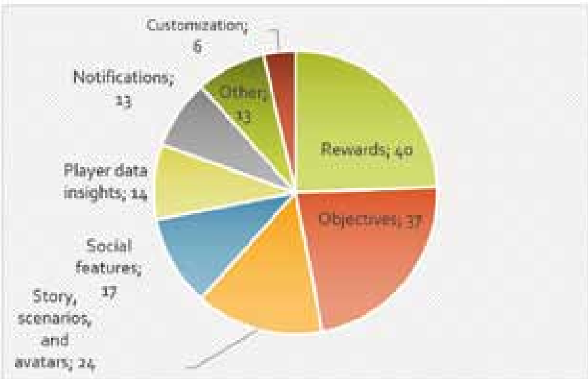



**Figure 2.20:** Game characteristics used by all empowerment methods.

For all empowerment methods, rewards were the most used game mechanic. It is most used for supporting behaviour, namely in 86% of the applications. Whereas for training skills and transferring knowledge, rewards are used in just over 60% of the applications. Training skills and transferring knowledge further have the same most popular mechanics, but the order and percentages of applications in which it is used differ. Supporting behaviour has different mechanics, while it also uses objectives in many of the applications (81%), it also uses social elements in more than half (57%) of the applications, and notifications and player data insights in 48% of the applications.

### 2.5.3 Evaluation studies and results

In this section, it is studied how the serious games and gamified interventions are evaluated and what kind of results are found.

#### 2.5.3.1 Evaluation studies and results for training skills

Most of the reviewed papers within this empowerment method, concerned a study [64, 48, 27, 12, 37, 67, 69, 52, 35, 83, 14, 49, 65, 57, 11, 62, 74, 41, 7, 51]. Only two papers [5, 22] describe a protocol or design. Of all these studies, only two studies had a duration of six months or more. One has a follow-up after 6 months [51], and the other has the last follow-up after 18 months [64]. This second study is also the study with the longest duration found in this review. Noticeable is that the protocol for this study has previously been described in another paper.

Most of the studies use an experimental setup [64, 27, 12, 35, 14, 41, 7], mostly consisting of two different conditions. Sometimes different versions of an intervention were compared with each other and a control group [12, 14]. In one of the experiments, healthy participants were compared with participants with OCD [41].

Moreover, effect oriented studies [48, 83, 11, 62] and effect oriented studied with multiple interactions [37, 69, 65, 57, 51], were also found in the review of this empowerment method. In one of the effect studies, three different version of the game were compared with each other [62], another study compared three different target devices [11]. However, both studies did not use a control condition to compare with, and it is therefore not seen as an experiment.

Finally, focus groups were used [67, 52, 49, 74]. Focus groups were either done with experts or professionals [74], or potential end users [52], or a combination of both [49]. One of the studies had four different evaluation stages, in which both potential end users and experts were used as participants [67].

The number of participants were very different. Most of the studies include between 11 and 100 participants (see Figure 2.21[5]), however 2 studies have >500 participants [64, 27]. Most of the participants are in the target group of the application, with some exceptions, for example one paper includes not only the target group (children) but also young adults [48]. Other previous examples are of focus groups with experts or professionals as participants.

When studying the sample sizes of the different types of studies (see Table 2.1[6]) it becomes clear that experiments have the highest average number of participants, but also a high standard deviation for the number of participants. Focus groups have, on average, the lowest number of participants. Both results are in accordance with the numbers one should expect based on the nature of the study type.

Most of the studies both study the effect of the intervention, as well as how players perceive it. Often, positive and/or promising results are found, although not always for all

[5]As one of the studies consisted of 4 different stages, for 3 of which the sample size was mentioned separately, these 3 studies are taken into account separately for this figure.

[6]For this table the same holds as for the previous figure, see Footnote 5

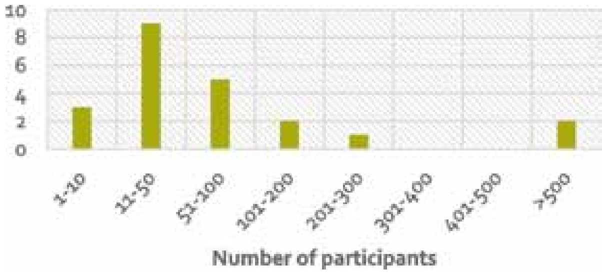



**Figure 2.21:** Number of participants in studies for training skills.

**Table 2.1:** Participants for different study types for training skills.

| | Minimum | Maximum | Average | Standard deviation |
|---|---|---|---|---|
| **Effect oriented** | 20 | 80 | 55.75 | 22.00 |
| **Effect oriented multiple interactions** | 11 | 155 | 58.2 | 51.51 |
| **Experimental** | 21 | 574 | 235.29 | 220.86 |
| **Focus groups** | 3 | 82 | 22.33 | 27.88 |

measurements. For example, in one of the studies it was found that the participants already had good knowledge about the topic of the game, and increases were only found for some measures [69]. The game that compared healthy participants with participants with OCD studied brain images to measure the effect of the intervention [41]. The longitudinal study found that results remained after six months, but not after 18 months [64].

#### 2.5.3.2 Evaluation studies and results for supporting behaviour

Slightly less than half of the papers reviewed within this empowerment method, describe a protocol for a study or design of an intervention [24, 84, 82, 47, 26, 29, 33, 66, 9]. Of the other papers, only one describes a longitudinal experimental study with a follow-up after one year [16], this study also describes a focus group as part of their research.[7]

Most of the studies are experiments [43, 16, 23, 80, 38, 79]. Two effect oriented studies were found: one in which there was only one interaction with the system [19], and one consisting of a usability test, and a 28-day pilot trial [36]. A high variety of participants in this category was found: while the usability test had 10 participants, and the pilot trial of the same intervention had 20 participants [36], the other effect study included 104 participants [19].

The remaining studies are focus group studies [16, 61, 77, 63, 13]. Some focus group used the System Usability Scale (SUS) as an evaluation method for the usability of the system [61, 77, 13]. One of the focus groups consisted of multiple trials, with different

[7]The number of participants for both studies different, and are used separately in the analysis of the participant numbers.

**Table 2.2:** Participants for different study types for supporting behaviour.

| | Minimum | Maximum | Average | Standard deviation |
|---|---|---|---|---|
| **Experimental** | 36 | 824 | 271.5 | 257.97 |
| **Focus groups** | 6 | 21 | 13.5 | 5.06 |

numbers of participants [77], these are taken into account separately for the following analyses.

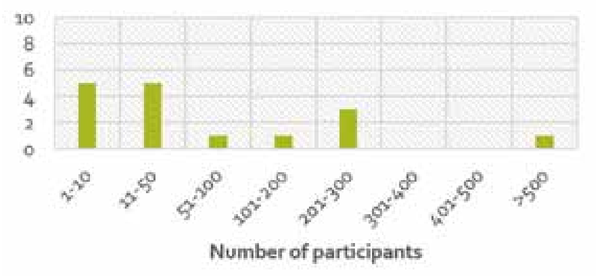



**Figure 2.22:** Number of participants in studies for supporting behaviour.

Overall, the most studies have less than 50 participants (see Figure 2.22). The highest number of participants is 824 for an experimental study with Dutch adolescents [43].

For the experimental studies and focus groups[8], see Table 2.2, this shows that the average of experimental studies is around 270, but with a high standard deviation, while the average of focus groups is much smaller, 14, and with a much lower standard deviation.

Most of the studies study the effect of the intervention and/or the experience of the users with the intervention. Besides this, one paper studied how the response time of a step with their system was related to the Berg Balance Score [19]. Another remarkable result was that of the study with the highest number of participants. It reported a low intervention completion, initially 2649 participants were included in the research but only 824 remained [43]. Moreover, this research only found an effect on the binge drinking behaviour of some subgroups in the study. One study found results that were negative and opposite of existing literature. It found that elderly explicitly refused gamification elements in a physical training programme [63].

#### 2.5.3.3 Evaluation studies and results for transferring knowledge

Most of the studies describe a study [18, 28, 81, 60, 4, 71, 25, 72, 34, 17], only three study protocols or designs of games were found in this category [10, 3, 78]. Only one of the studies uses data from at least half a year [18].

[8] As there were only two effect oriented studies, with a big difference in the number of participants, there are not taken into consideration here.

**Table 2.3:** Participants for different study types for transferring knowledge.

| | Minimum | Maximum | Average | Standard deviation |
|---|---|---|---|---|
| **Effect oriented** | 65 | 354 | 193.75 | 126.46 |
| **Experimental** | 45 | 157 | 95.5 | 45.26 |
| **Focus group** | 25 | 74 | 39.4 | 18.29 |

Four of the studies are effect oriented studies [18, 28, 81, 60] and four have an experimental setup [4, 71, 72, 34]. Of the effect oriented studies, most measure the effect after one interaction with their intervention [28, 81, 60]. For the development of ‘Pure Rush’ there were three stages, with different numbers of participants: the formative consultation (n = 115)[9], feasibility and acceptability testing (n = 25) and the final evaluation (n = 281). The feasibility and acceptability phase was in the form of a focus group, the final evaluation is an effect study [81]. The only effect oriented study with more than one interaction moment was the longitudinal study; which also consists of different phases (exploratory, implementation, administration, and evaluation). In which again also a focus group was incorporated [18]. One of the studies with an experimental setup also includes a focus group with a subset of the participants [72]. Other focus groups are described in [25, 17].

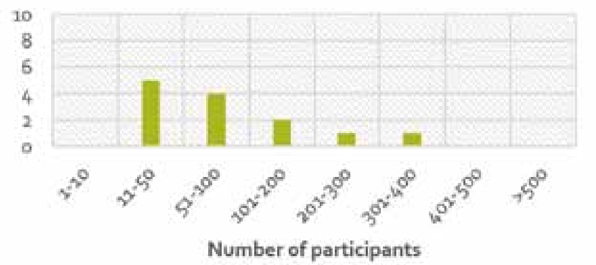



**Figure 2.23:** Number of participants in studies for transferring knowledge.

Figure 2.23 shows that most studies have 11–100 participants. Most of these participants are also the target group of the application. Table 2.3 shows that the effect oriented studies, on average, have the highest number of participants, and the focus groups have the smallest. Again, comparable to previous discussed empowerment methods, the focus groups also have a small standard deviation compared to the other categories.

Most studies focus on the effect of the intervention or how players perceived it. Often positive results are found and/or results can be used to further improve or extend the applications. For example, one study used a focus group approach and learned that while their game was educational and somewhat entertaining, the game was lacking real-life scenarios and player control [25].

[9] As this is not part of the evaluation of the game itself, this phase is not taken into account for Figure 2.24

#### 2.5.3.4 Evaluation studies and results for other interventions

The first intervention is a study with a focus group setup with 81 participants [45]. The participants, children from the same age group as the target group, were asked to draw for example characters, to study what storylines and characters could be used in the game.

The second intervention was evaluated with two experiments in which different conditions were compared, as well as a control condition [6]. The first experiment had 92 participants, the second experiment 120. However, it concerned healthy student participants, while the target group is people with acrophobia. No significant differences were found for the two conditions of the intervention that were compared.

#### 2.5.3.5 Conclusions about evaluation studies and results

Figure 2.24 shows that 24% of the reviewed papers describes a protocol or design of a game. For the empowerment method of supporting behaviour, most of the reviewed papers concerned such a protocol or design, while for other empowerment method, most papers described a study.

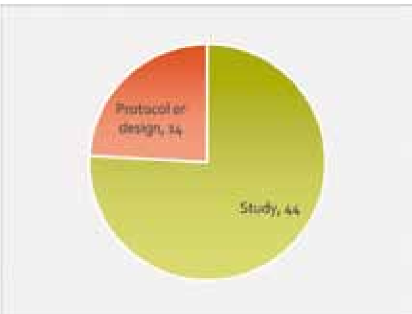



**Figure 2.24:** Ratio papers describing studies and protocols.

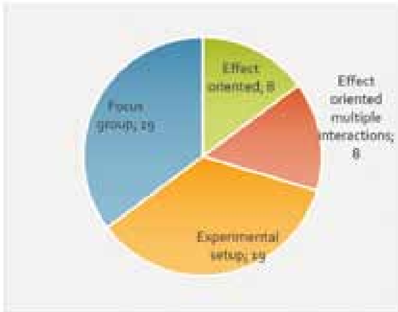



**Figure 2.25:** Ratio of study types.

Figure 2.25 shows that most of the studies are focus groups or studies with an experimental setup. It was found that focus groups are sometimes used in combination with other types of studies.

Figure 2.26 shows how many participants are used in all the studies. This shows that most studies have between 11 and 50 participants, followed by studies with 51–100 participants. The empowerment method of supporting behaviour is an exception to this, with the most participants between 1 and 50.

Experimental setups and effect oriented studies have the largest numbers of participants, but they also have a high standard deviation. Focus groups on the other hand often have a lower number of participants (<100, often <50), but also with a lower standard deviation.

The differences between the empowerment methods are small. Remarkable is that for supporting behaviour, most of the reviewed papers describe a protocol or design. Most

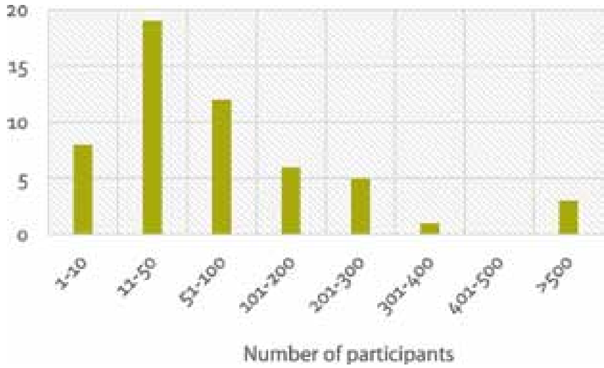



**Figure 2.26:** Number of participants in all studies.

of the studies are experiments, which could be related to the fact that there are many protocols. However, the number of participants for most of the studies is, on average, lower for this empowerment method, while it could be expected that, given the number of experimental setups, larger numbers of participants would have been used. However, it seems that there is a high variety in the sample sizes of the experimental studies.

The results of studies are often oriented on improving the current intervention, or measuring the effect. Sometimes specific aspects, such as target device or the affiliation of the target group with the game characters and story. However, only four of the studies contained data from more than 6 months. This, together with the fact that many results are very specific for the evaluated game (as they describe possible improvements), makes the generalizability of the results questionable. Moreover, the study with the follow-up after 18 months showed that results did not hold for so many months. As many of the studies lack such longitude results, it is unclear whether reported results will hold in the future. Moreover, the measurements used to evaluate a system are often specific for the system itself, this makes comparing and generalizing results also harder. In some papers it was found that the SUS was used for measuring the usability, such measurement methods make it possible to compare results from different studies with each other.

## 2.6 Conclusions

### 2.6.1 Summary and findings

The current chapter gives an overview of the domain of serious games and gamification to empower vulnerable target groups. In the introduction of this chapter the concepts vulnerability and empowering were explained. Next, this chapter explained how large target groups and goals of applications within the domain of serious gaming and gamification can be distinguished from each other by studying their objectives. With a classification based on different objectives, it is explained how applications differ from each other on their goal and target group, taking into account the context in which they are used. This classification explains how the specific objective that is studied in this review, differs from other serious games and gamified applications.

To give an overview of the domain of serious games and gamification with the objective to empower vulnerable target groups, the first part of this structured review focused on three questions:

- Which vulnerable target groups are studied?
- How can the target group's risk(s) be characterized?
- In what way do the interventions empower the target group?

Five different categories of target groups are found within the review, namely people vulnerable due to their (ordered on size): age, physical health condition, mental health condition, cultural background or socio-economic status. Age is the biggest target group, making 44% of all the papers.

Three large categories of how vulnerability risk are identified have been found within this review: health, safety and social. 80% of the vulnerabilities are related to health risks. However, different meanings for this category were found. For most target groups, their health risk is of a preventive nature, meaning that the application aims to prevent a future health risk. However, when it comes to people that are already having a specific health condition, the application most often aims to reduce the risks of that health situation, rather than prevent another health risk.

When studying how applications aim to empower their users, three main methods are found (ordered by size): training a skill, supporting behaviour, and transferring knowledge. For supporting behaviour, a division was found between applications that support physical activity, such as exergames, on the other hand applications were found that support so called lifestyle behaviours.

The age target group is the most diverse group: it has the most different types of risks and empowerment methods. Health risks, for the different target groups, are addressed with all types of empowerment methods, but mainly with behaviour support. Moreover, behaviour support is almost exclusively used for health risks. Training skills and transferring knowledge is mainly used for the other two vulnerability risks. Furthermore, it is found that for people with a physical health condition, only health risks are addressed. For people with a mental health condition this is almost the same, with one exception.

In this first part of the research it becomes clear that interventions often aim for vulnerabilities that are closely related to the characteristics of the vulnerable group, and that are often occurring, such as substance abuse among adolescents or the health risks of a health condition. More indirect vulnerabilities, such as social vulnerabilities for people with a specific health condition are rarely found. It seems that pathogenic vulnerabilities (situational vulnerabilities that are caused by injurious social phenomena) are not targeted by the current research projects. This is a gap in the current research, that is very interesting to study in the future.

In the second part of the review, the technologies, game mechanics and evaluation methods and results have been studied. We did not differentiate between serious games and gamification, as the descriptions of the interventions differ from paper to paper, it can be sometimes challenging to distinguish between these two concepts. We therefore decided to not make this distinction, also to avoid mistakes.

Concerning the platforms that are used it is shown that mobile and web-based applications are mostly used, but also desktop applications or other technologies such Wii Balance boards and Kinect are used.

When studying the game mechanics and characteristics, it became clear that they can be summarized in different categories, such as objectives, rewards, and story elements. While in each empowerment method it is clear that objectives and rewards play a major role, the different methods have a different preference for these mechanics.

Skill training uses rewards in 64% of the applications, and objectives in 59% of the applications. Story elements are also used in just over half of the applications (55%). For behaviour supporting applications, rewards are a very important component, 86% of the applications use this. This is again followed by objectives (81%). Stories are not often used for this empowerment method, but social elements (57%) and notifications and player data insights (48%) are. In knowledge transferring games, again rewards are the most used mechanic (62%). In the other empowerment methods, the second most used mechanic is objectives, but for knowledge transferring this is the story mechanic (46%). In the other methods, objectives are used in more than half of the applications, but for knowledge transferring this is only 38%.

Finally, the evaluation methods and results were analysed. 24% of the reviewed papers did not describe any results, but describe a study protocol or game design. In all empowerment methods but one, the majority of the papers describes a study. Only for the empowerment method of supporting behaviour the majority of the papers is a protocol or design. Of the studies, most studies have an experimental setup or focus group setup (35%), other studies measure the effect of an intervention after one or more interactions with the intervention (30%). Notable is that only four out of the 54 studies collected data over a period of six months or more. In general, studies have 11–100 participants. Focus groups often have a smaller number of participants (on average around 30), while experimental studies have a high number of participants (on average around 200). However, the standard deviation for the average number of participants for experimental studies guide high; while some there are three studies with more than 500 participants, there are also five studies with less than 50 participants. For focus groups this diversity is less big. Most focus groups contain at most 20 participants.

The results of the studies are mostly about the effect of the game, or about the attitude of participants towards the game. Often, these results are used to as preliminary results about the effect and/or to further improve the intervention. However, as most studies do not include a longitude follow-up, it is unclear how these effects hold. Moreover, the longest study showed that their effects did not hold after a certain amount of time [64]. Due to the lack of data for this amount of time, most studies cannot say anything about this.

### 2.6.2 Recommendations

It is often stated in reviews that papers in the domain of serious gaming and gamification lack a longitudinal and large effect study. Based on the performed literature review, a more specific observation can be made. While it is indeed a problem that studies often lack a large and long effect study, a more notable problem is the lack of generalizability of the research results. This might not only be a problem of the studies itself, but also of the descriptions of the applications. When looking at the descriptions of serious games and gamified applications, the descriptions of the games are very different from each other. While some papers describe their application with a high level of detail and by using commonly used terms for game mechanics, other games are described in a more story-like description in which the actual game mechanics might be less clear or make it harder to understand the role of certain game mechanics. In Section 2.5.2.2 and Section 2.5.2.3 two good examples of a description were mentioned ([16, 56]), that clearly explained which components are used in the game, and, in the case of the game from Section 2.5.2.2, how they are connected to behaviour change techniques [16]. Another good example is shown in the paper about AllyQuest [36], in which the game mechanics and scientific reasoning behind them is schematically shown.



Other domains, such as medicine, usually describe their studies in a very high detail, explaining a lot of factors so that others can reproduce the same study. Such an approach could also be beneficial for the domain of serious games and gamification. However, in addition to explaining how the study is setup (which is already done in papers with a study protocol), it is also important to explain the used components, their functioning and the scientific reasoning behind them. As the components of a serious game or gamified application are important for knowing how to reproduce a comparable intervention and subsequently compare studies.

However, describing games in such detail might be difficult. One reason for this is that game mechanics can have different meanings in different contexts. When assigning points to exercises in a game with a leaderboard, this might have a completely different effect compared to using points in a game in which users unlock rewards based on their points. Moreover, aspects that were not meant as game mechanics might turn out to influence the behaviour of users. For example, if you have a wardrobe with unlockable clothes for an avatar, users might not be motivated by the fact that they can customize their avatar, but instead be motivated by the fact that they want to unlock every item from the wardrobe. This can also be influenced by personal preferences of players. Measuring the effect of individual game mechanics is often not performed in studies.

This review identified different labels to characterize serious games and gamified applications; this is done for the target groups, vulnerability risks, empowerment methods and game mechanics. As the domain of serious gaming and gamification is very broad (in the sense of target groups and vulnerabilities), there are different disciplines involved in the different projects. Therefore, it might be challenging to find one universal way to describe games (and their mechanics) in a more standardized way. However, doing so, such as is attempted in this review, makes the field more transparent which can lead to more generalizable results and thereby more effective serious games and gamified applications. No previous attempts for this specific domain of empowering vulnerable target groups have been found.

Finally, what also contributes to the lack of generalizability is that the measuring instruments used in the evaluations of the serious games and gamified application are often developed for that specific purpose. By using more frequently used measures, such as the SUS for usability, it is possible to compare the outcomes of different studies.

In conclusion, this chapter has contributed to showing the diversity in a promising field of serious gaming and gamification. As this review has shown, it is possible to apply serious gaming and gamification to many different behaviours, which is often positively evaluated by the potential users and has positive effects. However, this review has also shown that the field lacks a generalizability of the described results, which we believe is due to the lack of standardized ways to describe serious games and gamification. For future research it is interesting to apply labels to describe different types or genres of games, in order to make it easier to find clusters of games that are somewhat comparable. More advanced would be to move towards a more standardized way to describe games in research papers, or at least identify the most important game mechanics that form the core of an application. Moreover, making sure that the described intervention and the code publically available, would also be beneficial for this research field as is allows other researches from outside the project to reuse existing material.

## 2.A Overview of reviewed papers and their classifications

| **Description** | **Target group** | **Type of risk** | **Empowerment methods** | **Technology** | **Game characteristics and mechanics** | **Study characteristics** | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | **Type of study** | **Duration** | **Number of participants** |
| Alcohol alert: game to reduce binge drinking among adolescents [43] | Age | Health | Supporting behaviour | Web (played on computer) | Objectives; Rewards; Story, scenarios & avatars; Notifications | Experimental | Short | >500 |
| SPARX-R: preventing depression for adolescents [64] | Age | Health | Training skills | Web (played on school computer); Other (paper notebook) | Objectives | Experimental | Long | >500 |
| Learning game for training child bicyclists' situation awareness [48] | Age | Safety | Training skills | Mobile | Rewards | Effect | Short | 51–100 |
| Using virtual reality in computer classes to reduce loneliness among elderly [5] | Age | Social | Training skills | Desktop (VR) | Unspecified | Protocol | | |
| Serious game about food safety for young customers [18] | Age | Safety | Transferring knowledge | Web | Objectives; Rewards; Story, scenarios & avatars; Other | Effect multiple interactions; Focus group | Long | 301–400; 51–100 |
| Happy 8-12: Emotional education program for the assertive resolution of conflicts among peers (adolescents) [27] | Age | Social | Training skills | Desktop | Story, scenarios & avatars | Experimental | Short | >500 |

| Description | Target group | Type of risk | Empowerment methods | Technology | Game characteristics and mechanics | Study characteristics | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Type of study | Duration | Number of participants |
| Gnam's Planet: intervention for healthy lifestyle promotion in adolescents [28] | Age | Health | Transferring knowledge | Web (played on a computer) | Objectives; Rewards | Effect | Short | 51–100 |
| Step Smart Challenge: gamified intervention for encouraging physical activity in adolescents [16] | Age | Health | Supporting behaviour | Web (with workbook) | Objectives; Rewards; Social; | Experimental; Focus group | Long | 201–300; 11–50 |
| CampusGANDR: college alcohol intervention [23] | Age | Health | Supporting behaviour | Mobile | Objectives; Rewards; Social | Experimental | Short | 201–300 |
| Pure Rush: an online serious game for drug education for adolescents [81] | Age | Health | Transferring knowledge | Mobile; Web | Rewards; Story; Customization | Effect; Focus group; | Short | 201–300; 11–50 |
| The Fling: serious game to train behavioural control in adolescents related to alcohol usage [12] | Age | Health | Training skills | Desktop | Objectives; Rewards; Story, scenarios & avatars | Experimental | Short | 101–200 |
| Active Team: app-based physical activity intervention for inactive adults [24] | Age | Health | Supporting behaviour | Mobile | Objectives; Rewards; Player data insights; Notifications; Social | Protocol | | |
| Cognitive Function Training System Using Game-Based Design for Elderly Drivers [37] | Age | Safety | Training skills | Mobile | Objectives; Rewards | Effect multiple interactions | Short | 11–50 |
| Healthy teens @ school: online program for promoting a healthy lifestyle and reducing the risk of eating disorders and obesity for adolescents [61] | Age | Health | Supporting behaviour | Web | Objectives; Rewards; Player data insights | Focus group | Short | 1–10 |
| Strength and balance exergames to reduce falls risk for elderly [80] | Age | Health | Supporting behaviour | Other (Kinect) | Unspecified | Experimental | Short | 51–100 |
| Online educational program using game-based learning to improve nutrition and physical activity for students [10] | Age | Health | Transferring knowledge | Web | Rewards; Notifications; Social; Other | Protocol | | |
| FallSensing Games: fall prevention multiplayer game for senior care centres [77] | Age | Health | Supporting behaviour | Other (wearable sensor + tv) | Objectives; Rewards; Social; Other | Focus group | Short | 2x 1–10; 11–50 |
| Train and Win: digital home-based physical training program for elderly [63] | Age | Health | Supporting behaviour | Other (Kinect) | Rewards; Story, scenarios & avatars | Focus group | Short | 11–50 |

| **Description** | **Target group** | **Type of risk** | **Empowerment methods** | **Technology** | **Game characteristics and mechanics** | **Study characteristics** | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | **Type of study** | **Duration** | **Number of participants** |
| Space Adventures: a serious game for childhood obesity prevention [60] | Age | Health | Transferring knowledge | Desktop | Story, scenarios & avatars | Effect | Short | 51–100 |
| A serious game-based solution to prevent bullying for children [67] | Age | Social | Training skills | Mobile; Desktop | Objectives; Rewards; Story, scenarios & avatars | Focus group | Short | 1–10; 11–50; 51–100 |
| Gamified e-learning platform to improve traffic safety among elementary school pupils [69] | Age | Safety | Training skills | Desktop | Objectives; Rewards; Player data insights; Story, scenarios & avatars; Other | Effect multiple interactions | Short | 11–50 |
| Trucs tegen babbeltrucs: A serious game for training verbal resilience to doorstep scams [52] | Age | Safety | Training skills | Mobile | Rewards; Story, scenarios & avatars | Focus group | Short | 1–10 |
| A stepping game for older adults [19] | Age | Health | Supporting behaviour | Other (Wii Balance Board) | Objectives; Rewards | Effect | Short | 101–200 |
| PlayForward: Elm City Stories: public health videogame intervention about HIV for adolescents [35] | Age | Health | Training skills | Unspecified | Objectives; Story, scenarios & avatars; Other | Experimental | Short | 201–300 |
| GAMETEEN SYSTEM: a VR-based serious game to regulate joy in adolescents [83] | Age | Health | Training skills | Mobile; Desktop; Other (RGB-D camera) | Objectives; Rewards; Other | Effect | Short | 51–100 |
| A gamified application for assessment of balance and fall prevention for elderly [84] | Age | Health | Supporting behaviour | Mobile (with non-invasive wearable sensor) | Unspecified | Protocol | | |
| ONESELF: web-based intervention for rheumatoid arthritis patients [4] | Physical health condition | Health | Transferring knowledge | Web | Rewards; Social; Other | Experimental | Short | 101–200 |
| An oral health education video game for high caries risk children [3] | Physical health condition | Health | Transferring knowledge | Mobile | Unspecified | Protocol | | |
| GoFar: reducing disruptive behaviour in children with Fetal Alcohol Spectrum Disorders [14] | Physical health condition | Health | Training skills | Desktop | Objectives; Rewards | Experimental | Short | 11–50 |
| Smartphone application for weight loss in overweight and obese adolescents [82] | Physical health condition | Health | Supporting behaviour | Mobile | Objectives; Rewards; Player data insights; Customization | Protocol | | |

| Description | Target group | Type of risk | Empowerment methods | Technology | Game characteristics and mechanics | Study characteristics | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Type of study | Duration | Number of participants |
| AllyQuest: smartphone application to support engagement in care and medication adherence for HIV-positive young men [36] | Physical health condition | Health | Supporting behaviour | Mobile | Objectives; Rewards; Player data insights; Story, scenarios & avatars; Notifications; Social | Effect multiple interactions | Short | 1–10; 11–50 |
| Serious game for older adults undergoing treatment for cancer [49] | Physical health condition | Health | Training skills | Unspecified | Story, scenarios & avatars | Focus group | Short | 11–50 |
| Epic Allies: a gamified mobile phone app to improve engagement in care, antiretroviral uptake, and adherence among HIV positive [47] | Physical health condition | Health | Supporting behaviour | Unspecified | Objectives; Rewards; Player data insights; Notifications; Social; Customization | Protocol | | |
| PROVITAO: gamified educational programme for childhood obesity [71] | Physical health condition | Health | Transferring knowledge | Mobile; Other (variety of games, also commercial games like Wii Fit Plus) | Objectives; Rewards; Player data insights; Other | Experimental | Short | 11–50 |
| INTERACCT: serious game for health of children after cancer treatment [45] | Physical health condition | Health | Other | Web (with Kinect or Android data) | Objectives; Social; Other; | Focus group | Short | 51–100 |
| Exercise intervention program using social incentives and gamification for obese children [26] | Physical health condition | Health | Supporting behaviour | Mobile | Rewards; Social | Protocol | | |
| MyHeartMate: a game-based app to promote behaviour change in patients with cardiovascular disease [29] | Physical health condition | Health | Supporting behaviour | Mobile (with wearable) | Objectives; Rewards; Player data insights; Story, scenarios & avatars; Notifications; Social; Other | Protocol | | |
| STEP UP: social incentives to encourage physical activity and understand predictors for overweight and obese adults [33] | Physical health condition | Health | Supporting behaviour | Web (with wearable device) | Objectives; Rewards; Notifications; Social | Protocol | | |
| Mission: Schweiniehund: behaviour change technique-based smartphone game to improve intrinsic motivation and physical activity adherence in patients with type 2 diabetes [38] | Physical health condition | Health | Supporting behaviour | Mobile | Objectives; Rewards; Player data insights; Notifications; Social; Customization | Experimental | Short | 11–50 |

| Description | Target group | Type of risk | Empowerment methods | Technology | Game characteristics and mechanics | Study characteristics | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | **Type of study** | **Duration** | **Number of participants** |
| SIGMA: an evidence-based gamified mHealth intervention for overweight young adults with maladaptive eating habits [66] | Physical health condition | Health | Supporting behaviour | Mobile | Objectives; Rewards; Player data insights; Story, scenarios & avatars; Notifications | Protocol | | |
| A holistic technology-based solution for prevention and management of diabetic foot complications [13] | Physical health condition | Health | Supporting behaviour | Mobile (with thermal camera) | Objectives; Rewards; Player data insights; Notifications | Focus group | Short | 1–10 |
| Virtual Coach: intelligent agent and virtual game to support education in e-health for children with type 1 diabetes [65] | Physical health condition | Health | Training skills | Unspecified | Rewards; Story, scenarios & avatars | Effect multiple interactions | Short | 11–50 |
| 0Phobia: serious game for patients with acrophobia [22] | Mental health condition | Health | Training skills | Mobile (with VR goggles) | Objectives; Rewards; Player data insights; Story, scenarios & avatars; Notifications | Protocol | | |
| Trauma- Gameplay: game for people with PTSD to reduce intrusions [6] | Mental health condition | Health | Other | Unspecified | Objectives | Experimental | Short | 51–100; 101–200 |
| Take Control: virtual reality cue refusal video game for alcohol and cigarette recovery support [57] | Mental health condition | Health | Training skills | Other (Kinect) | Rewards | Effect multiple interactions | Short | 51–100 |
| Trash that cigarette: an avatar-based anti-smoking game on smoking cessation intent [11] | Mental health condition | Health | Training skills | Mobile | Objectives; Story, scenarios & avatars | Effect | Short | 51–100 |
| OnTrack› TheGame: computer-based role-playing game for young people with psychosis [62] | Mental health condition | Health | Training skills | Web | Rewards; Story, scenarios & avatars; Customization | Effect | Short | 11–50 |
| Let's find letters: learning aid for dyslexic children [74] | Mental health condition | Health | Training skills | Unspecified | Objectives; Rewards | Focus group | Short | 1–10 |
| RAW HAND: mobile serious game in the treatment of OCD [41] | Mental health condition | Health | Training skills | Mobile | Other | Experimental | Short | 11–50 |

| Description | Target group | Type of risk | Empowerment methods | Technology | Game characteristics and mechanics | Study characteristics | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Type of study | Duration | Number of participants |
| IntegraGame: vocational training tool for persons with intellectual disability [7] | Mental health condition | Social | Training skills | Unspecified | Objectives | Experimental | Short | 11–50 |
| Fast Car: Web-based HIV prevention game for rural adolescents [25] | Cultural background | Health | Transferring knowledge | Web | Rewards | Focus group | Short | 11–50 |
| Making smart choices: a serious game for sex education for young adolescents [46] | Cultural background | Health | Transferring knowledge | | | | | |
| None in Three: violence prevention game for young people in the Caribbean region [78] | Cultural background | Safety | Transferring knowledge | Desktop | Objectives; Story, scenarios & avatars | Protocol | | |
| Tumaini: A smartphone game to prevent HIV among young Africans [72] | Cultural background | Health | Transferring knowledge | Mobile | Story, scenarios & avatars; Customization; Other | Experimental; Focus group | Short | 11–50; 51–100 |
| Improving sexual health education programs for adolescent (in Sub Saharan Africa) through game-based learning and gamification [34] | Cultural background | Health | Transferring knowledge | Desktop (at school) | Rewards; Story, scenarios & avatars; Social; Other; | Experimental | Short | 101–200 |
| PROCEE: a PROstate Cancer Evaluation and Education serious game for African Caribbean men [17] | Cultural background | Health | Transferring knowledge | Unspecified (only mentions tablet used for focus group) | Objectives; Player data insights; Story, scenarios & avatars | Focus group | Short | 11–50 |
| Online evidence-based parenting program for vulnerable parents [51] | Socio-economic status | Social | Training skills | Mobile; Desktop | Objectives; Rewards; Story, scenarios & avatars; Notifications; Social; Other; | Effect multiple interactions | Long | 101–200 |
| Balance it: game intervention for overweight-related behaviours for adolescents [79] | Socio-economic status | Health | Supporting behaviour | Mobile | Objectives; Rewards; Notifications; Social | Experimental | Short | 201–300 |
| BUZZING: a serious game for unemployed young people their transition into employment [9] | Socio-economic status | Social | Supporting behaviour | Mobile; Web | Objectives; Player data insights; Social | Protocol | | |

# Bibliography

[1] Empowerment - Cambridge Dictionary [online]. https://dictionary.cambridge.org/dictionary/english/empowerment

[2] Abt, C.C.: Serious games. University press of America (1987)

[3] Aljafari, A., Rice, C., Gallagher, J.E., Hosey, M.T.: An oral health education video game for high caries risk children: study protocol for a randomized controlled trial. Trials **16**(1), 1–10 (2015). https://doi.org/10.1186/s13063-015-0754-6

[4] Allam, A., Kostova, Z., Nakamoto, K., Schulz, P.J.: The effect of social support features and gamification on a web-based intervention for rheumatoid arthritis patients: Randomized controlled trial. J Med Internet Res **17**(1), e14 (Jan 2015). https://doi.org/10.2196/jmir.3510

[5] Antunes, T.P.C., de Oliveira, A.S.B., Crocetta, T.B., de Lima Antão, J.Y.F., de Almeida Barbosa, R.T., Guarnieri, R., Massetti, T., de Mello Monteiro, C.B., de Abreu, L.C.: Computer classes and games in virtual reality environment to reduce loneliness among students of an elderly reference center: study protocol for a randomised cross-over design. Medicine **96**(10) (2017). https://doi.org/10.1097/MD.0000000000005954

[6] Asselbergs, J., Sijbrandij, M., Hoogendoorn, E., Cuijpers, P., Olie, L., Oved, K., Merkies, J., Plooijer, T., Eltink, S., Riper, H.: Development and testing of traumagameplay: an iterative experimental approach using the trauma film paradigm. European Journal of Psychotraumatology **9**(1), 1424447 (2018). https://doi.org/10.1080/20008198.2018.1424447, pMID: 29441151

[7] von Barnekow, A., Bonet-Codina, N., Tost, D.: Can 3d gamified simulations be valid vocational training tools for persons with intellectual disability? Methods of information in medicine **56**(02), 162–170 (2017). https://doi.org/10.3414/ME16-02-0014

[8] Barwick, J., Watkins, D., Kirk, E., Law, E.: Adventures with lex: The gamification of research? Convergence **24**(3), 229–250 (2018). https://doi.org/10.1177/1354856516677682

[9] Beaton, S.: Buzzing—a theory-based impact evaluation design. Evaluation Journal of Australasia **16**(4), 21–29 (2016). https://doi.org/10.1177/1035719X1601600404

[10] Belogianni, K., Ooms, A., Ahmed, H., Nikoletou, D., Grant, R., Makris, D., Moir, H.J.: Rationale and design of an online educational program using game-based learning to improve nutrition and physical activity outcomes among university students in the united kingdom. Journal of the American College of Nutrition **38**(1), 23–30 (2019). https://doi.org/10.1080/07315724.2018.1476929, pMID: 30071183

[11] bMiami, V.: Effects of an avatar-based anti-smoking game on smoking cessation intent. Medicine Meets Virtual Reality 22: NextMed/MMVR22 **220**, 15 (2016)

[12] Boendermaker, W.J., Veltkamp, R.C., Peeters, M.: Training behavioral control in adolescents using a serious game. Games for Health Journal **6**(6), 351–357 (2017). https://doi.org/10.1089/g4h.2017.0071, pMID: 28926286

[13] Cleland, I., Rafferty, J., Synnott, J., Cundell, J., Boyd, A., Nugent, C., Chaurasia, P., Morrison, G., Madill, G., Monk-Ozgul, L., et al.: A holistic technology-based solution for prevention and management of diabetic foot complications. In: International Conference on Ubiquitous Computing and Ambient Intelligence. pp. 453–465. Springer (2017). https://doi.org/10.1007/978-

3-319-67585-5_46

[14] Coles, C.D., Kable, J.A., Taddeo, E., Strickland, D.C.: A metacognitive strategy for reducing disruptive behavior in children with fetal alcohol spectrum disorders: Go-far pilot. Alcoholism: Clinical and Experimental Research **39**(11), 2224–2233 (2015). https://doi.org/https://doi.org/10.1111/acer.12885

[15] Conaway, R., Garay, M.C.: Gamification and service marketing. SpringerPlus **3**(1), 1–11 (2014). https://doi.org/10.1186/2193-1801-3-653

[16] Corepal, R., Best, P., O'Neill, R., Tully, M.A., Edwards, M., Jago, R., Miller, S.J., Kee, F., Hunter, R.F.: Exploring the use of a gamified intervention for encouraging physical activity in adolescents: a qualitative longitudinal study in northern ireland. BMJ Open **8**(4) (2018). https://doi.org/10.1136/bmjopen-2017-019663

[17] Cosma, G., Brown, D., Shopland, N., Battersby, S., Seymour-Smith, S., Archer, M., Khan, M., Pockley, A.G.: Procee: a prostate cancer evaluation and education serious game for african caribbean men. Journal of Assistive Technologies (2016). https://doi.org/10.1108/JAT-12-2015-0035

[18] Crovato, S., Pinto, A., Giardullo, P., Mascarello, G., Neresini, F., Ravarotto, L.: Food safety and young consumers: Testing a serious game as a risk communication tool. Food Control **62**, 134–141 (2016). https://doi.org/https://doi.org/10.1016/j.foodcont.2015.10.009

[19] Deacon, M., Parsons, J., Mathieson, S., Davies, T.C.: Can wii balance? evaluating a stepping game for older adults. IEEE Transactions on Neural Systems and Rehabilitation Engineering **26**(9), 1783–1793 (2018). https://doi.org/10.1109/TNSRE.2018.2862146

[20] Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: Defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. p. 9–15. MindTrek '11, Association for Computing Machinery, New York, NY, USA (2011). https://doi.org/10.1145/2181037.2181040

[21] Dicheva, D., Dichev, C., Agre, G., Angelova, G.: Gamification in education: A systematic mapping study. Journal of Educational Technology & Society **18**(3), 75–88 (2015)

[22] Donker, T., Van Esveld, S., Fischer, N., Van Straten, A.: 0phobia–towards a virtual cure for acrophobia: study protocol for a randomized controlled trial. Trials **19**(1), 1–11 (2018). https://doi.org/10.1186/s13063-018-2704-6

[23] Earle, A.M., LaBrie, J.W., Boyle, S.C., Smith, D.: In pursuit of a self-sustaining college alcohol intervention: Deploying gamified pnf in the real world. Addictive Behaviors **80**, 71–81 (2018). https://doi.org/https://doi.org/10.1016/j.addbeh.2018.01.005

[24] Edney, S., Plotnikoff, R., Vandelanotte, C., Olds, T., De Bourdeaudhuij, I., Ryan, J., Maher, C.: "active team" a social and gamified app-based physical activity intervention: randomised controlled trial study protocol. BMC public health **17**(1), 1–10 (2017). https://doi.org/10.1186/s12889-017-4882-7

[25] Enah, C., Piper, K., Moneyham, L.: Qualitative evaluation of the relevance and acceptability of a web-based hiv prevention game for rural adolescents. Journal of Pediatric Nursing **30**(2), 321–328 (2015). https://doi.org/https://doi.org/10.1016/j.pedn.2014.09.004

[26] Fang, Y., Ma, Y., Mo, D., Zhang, S., Xiang, M., Zhang, Z.: Methodology of an exercise intervention program using social incentives and gamification for obese children. BMC public

health **19**(1), 1–10 (2019). https://doi.org/10.1186/s12889-019-6992-x

[27] Filella, G., Cabello, E., Pérez-Escoda, N., Ros-Morente, A., et al.: Evaluation of the emotional education program “happy 8-12” for the assertive resolution of conflicts among peers (2016). https://doi.org/10.14204/ejrep.40.15164

[28] Fraticelli, F., Marchetti, D., Polcini, F., Mohn, A.A., Chiarelli, F., Fulcheri, M., Vitacolonna, E.: Technology-based intervention for healthy lifestyle promotion in italian adolescence. Annali dell'Istituto superiore di sanita **52**(1), 123–127 (2016)

[29] Gallagher, R., Chow, C., Parker, H., Neubeck, L., Celermajer, D., Redfern, J., Tofler, G., Buckley, T., Schumacher, T., Ferry, C., Whitley, A., Chen, L., Figtree, G.: Design and rationale of the myheartmate study: a randomised controlled trial of a game-based app to promote behaviour change in patients with cardiovascular disease. BMJ Open **9**(5) (2019). https://doi.org/10.1136/bmjopen-2018-024269

[30] Gentry, S.V., Gauthier, A., Ehrstrom, B.L., Wortley, D., Lilienthal, A., Car, L.T., Dauwels-Okutsu, S., Nikolaou, C.K., Zary, N., Campbell, J., et al.: Serious gaming and gamification education in health professions: systematic review. Journal of medical Internet research **21**(3), e12994 (2019). https://doi.org/10.2196/12994

[31] Graafland, M., Schraagen, J.M., Schijven, M.P.: Systematic review of serious games for medical education and surgical skills training. British Journal of Surgery **99**(10), 1322–1330 (09 2012). https://doi.org/10.1002/bjs.8819

[32] Hamari, J., Koivisto, J., Sarsa, H.: Does gamification work? – a literature review of empirical studies on gamification. In: 2014 47th Hawaii International Conference on System Sciences. pp. 3025–3034 (2014). https://doi.org/10.1109/HICSS.2014.377

[33] Harrison, J.D., Jones, J.M., Small, D.S., Rareshide, C.A., Szwartz, G., Steier, D., Guszcza, J., Kalra, P., Torio, B., Reh, G., Hilbert, V., Patel, M.S.: Social incentives to encourage physical activity and understand predictors (step up): Design and rationale of a randomized trial among overweight and obese adults across the united states. Contemporary Clinical Trials **80**, 55–60 (2019). https://doi.org/https://doi.org/10.1016/j.cct.2019.04.001

[34] Haruna, H., Hu, X., Chu, S.K.W., Mellecker, R.R., Gabriel, G., Ndekao, P.S.: Improving sexual health education programs for adolescent students through game-based learning and gamification. International Journal of Environmental Research and Public Health **15**(9) (2018). https://doi.org/10.3390/ijerph15092027, `https://www.mdpi.com/1660-4601/15/9/2027`

[35] Hieftje, K., Pendergrass, T., Montanaro, E., Kyriakides, T., Florsheim, O., Fiellin, L.: “but do they like it?” participant satisfaction and gameplay experience of a public health videogame intervention in adolescents. In: 2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH). pp. 1–7 (2018). https://doi.org/10.1109/SeGAH.2018.8401349

[36] Hightow-Weidman, L., Muessig, K., Knudtson, K., Srivatsa, M., Lawrence, E., LeGrand, S., Hotten, A., Hosek, S.: A gamified smartphone app to support engagement in care and medication adherence for hiv-positive young men who have sex with men (allyquest): Development and pilot study. JMIR Public Health Surveill **4**(2), e34 (Apr 2018). https://doi.org/10.2196/publichealth.8923

[37] Hiraoka, T., Wang, T.W., Kawakami, H.: Cognitive function training system using

game-based design for elderly drivers. IFAC-PapersOnLine **49**(19), 579–584 (2016). https://doi.org/https://doi.org/10.1016/j.ifacol.2016.10.613, 13th IFAC Symposium on Analysis, Design, and Evaluation ofHuman-Machine Systems HMS 2016

[38] Höchsmann, C., Infanger, D., Klenk, C., Königstein, K., Walz, S.P., Schmidt-Trucksäss, A.: Effectiveness of a behavior change technique–based smartphone game to improve intrinsic motivation and physical activity adherence in patients with type 2 diabetes: Randomized controlled trial. JMIR Serious Games **7**(1), e11444 (Feb 2019). https://doi.org/10.2196/11444

[39] Höchsmann, C., Walz, S.P., Schäfer, J., Holopainen, J., Hanssen, H., Schmidt-Trucksäss, A.: Mobile exergaming for health—effects of a serious game application for smartphones on physical activity and exercise adherence in type 2 diabetes mellitus—study protocol for a randomized controlled trial. Trials **18**(1), 1–17 (2017). https://doi.org/10.1186/s13063-017-1853-3

[40] Hoffmaster, B.: What does vulnerability mean? Hastings Center Report **36**(2), 38–45 (2006). https://doi.org/10.1353/hcr.2006.0024 10.1353/hcr.2006.0024

[41] Hong, J.S., Kim, S.M., Aboujaoude, E., Han, D.H.: Investigation of a mobile "serious game" in the treatment of obsessive–compulsive disorder: A pilot study. Games for Health Journal **7**(5), 317–326 (2018). https://doi.org/10.1089/g4h.2017.0158, pMID: 30129775

[42] Hunicke, R., LeBlanc, M., Zubek, R.: Mda: A formal approach to game design and game research. In: Proceedings of the AAAI Workshop on Challenges in Game AI. vol. 4, p. 1722. San Jose, CA (2004)

[43] Jander, A., Crutzen, R., Mercken, L., Candel, M., de Vries, H.: Effects of a web-based computer-tailored game to reduce binge drinking among dutch adolescents: A cluster randomized controlled trial. J Med Internet Res **18**(2), e29 (Feb 2016). https://doi.org/10.2196/jmir.4708

[44] Johnson, D., Horton, E., Mulcahy, R., Foth, M.: Gamification and serious games within the domain of domestic energy consumption: A systematic review. Renewable and Sustainable Energy Reviews **73**, 249–264 (2017). https://doi.org/https://doi.org/10.1016/j.rser.2017.01.134

[45] Kayali, F., Peters, K., Kuczwara, J., Reithofer, A., Martinek, D., Wölfle, R., Mateus-Berr, R., Lehner, Z., Silbernagl, M., Sprung, M., et al.: Participatory game design for the interacct serious game for health. In: Joint International Conference on Serious Games. pp. 13–25. Springer (2015). https://doi.org/10.1007/978-3-319-19126-3_2

[46] Kwan, A.C., Chu, S.K., Hong, A.W., Tam, F., Lee, G.M., Mellecker, R.: Making smart choices: A serious game for sex education for young adolescents. International Journal of Game-Based Learning (IJGBL) **5**(1), 18–30 (2015). https://doi.org/10.4018/ijgbl.2015010102

[47] LeGrand, S., Muessig, K.E., Platt, A., Soni, K., Egger, J.R., Nwoko, N., McNulty, T., Hightow-Weidman, L.B.: Epic allies, a gamified mobile phone app to improve engagement in care, antiretroviral uptake, and adherence among young men who have sex with men and young transgender women who have sex with men: Protocol for a randomized controlled trial. JMIR Res Protoc **7**(4), e94 (Apr 2018). https://doi.org/10.2196/resprot.8811

[48] Lehtonen, E., Sahlberg, H., Rovamo, E., Summala, H.: Learning game for training child bicyclists' situation awareness. Accident Analysis & Prevention **105**, 72–83 (2017). https://doi.org/https://doi.org/10.1016/j.aap.2016.07.036, improving cyclist safety through sci-

entific research, ICSC2015.

[49] Loerzel, V., Clochesy, J., Geddie, P.: Using a community advisory board to develop a serious game for older adults undergoing treatment for cancer. Applied Nursing Research **39**, 207–210 (2018). https://doi.org/https://doi.org/10.1016/j.apnr.2017.11.030

[50] Loh, C.S., Sheng, Y., Ifenthaler, D.: Serious games analytics: Theoretical framework. In: Serious games analytics, pp. 3–29. Springer (2015). https://doi.org/10.1007/978-3-319-05834-4_1

[51] Love, S.M., Sanders, M.R., Turner, K.M., Maurange, M., Knott, T., Prinz, R., Metzler, C., Ainsworth, A.T.: Social media and gamification: Engaging vulnerable parents in an online evidence-based parenting program. Child Abuse & Neglect **53**, 95–107 (2016). https://doi.org/https://doi.org/10.1016/j.chiabu.2015.10.031, issues and Solutions in the Implementation of Evidence-informed Child Welfare Practices

[52] van der Lubbe, L.M., Gerritsen, C., Formolo, D., Otte, M., Bosse, T.: A serious game for training verbal resilience to doorstep scams. In: International Conference on Games and Learning Alliance. pp. 110–120. Springer (2018). https://doi.org/10.1007/978-3-030-11548-7_11

[53] Mackenzie, C., Rogers, W., Dodds, S.: Introduction: What is vulnerability and why does it matter for moral theory. Vulnerability: New essays in ethics and feminist philosophy pp. 1–29 (2014)

[54] Maher, C., Ferguson, M., Vandelanotte, C., Plotnikoff, R., De Bourdeaudhuij, I., Thomas, S., Nelson-Field, K., Olds, T.: A web-based, social networking physical activity intervention for insufficiently active adults delivered via facebook app: Randomized controlled trial. J Med Internet Res **17**(7), e174 (Jul 2015). https://doi.org/10.2196/jmir.4086

[55] Majuri, J., Koivisto, J., Hamari, J.: Gamification of education and learning: A review of empirical literature. In: Proceedings of the 2nd international GamiFIN conference, GamiFIN 2018. CEUR-WS (2018)

[56] Marchetti, D., Fraticelli, F., Polcini, F., Lato, R., Pintaudi, B., Nicolucci, A., Fulcheri, M., Mohn, A., Chiarelli, F., Di Vieste, G., Vitacolonna, E.: Preventing adolescents' diabesity: Design, development, and first evaluation of "gustavo in gnam's planet". Games for Health Journal **4**(5), 344–351 (2015). https://doi.org/10.1089/g4h.2014.0107, pMID: 26287925

[57] Metcalf, M., Rossie, K., Stokes, K., Tallman, C., Tanner, B.: Virtual reality cue refusal video game for alcohol and cigarette recovery support: Summative study. JMIR Serious Games **6**(2), e7 (Apr 2018). https://doi.org/10.2196/games.9231

[58] Michael, D.R., Chen, S.L.: Serious games: Games that educate, train, and inform. Muska & Lipman/Premier-Trade (2005)

[59] Morschheuser, B., Hamari, J., Koivisto, J.: Gamification in crowdsourcing: A review. In: 2016 49th Hawaii International Conference on System Sciences (HICSS). pp. 4375–4384 (2016). https://doi.org/10.1109/HICSS.2016.543

[60] Navarro, L.P., Ardila, D.P., Pires, M., Marques, J.: Space adventures: a serious game for childhood obesity prevention. In: VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th-28th, 2016. pp. 149–152. Springer (2017). https://doi.org/10.1007/978-981-10-4086-3_38

[61] Nitsch, M., Adamcik, T., Kuso, S., Zeiler, M., Waldherr, K.: Usability and engagement evaluation of an unguided online program for promoting a healthy lifestyle and reducing the risk for eating disorders and obesity in the school setting. Nutrients **11**(4) (2019). https://doi.org/10.3390/nu11040713

[62] Olivet, J., Haselden, M., Piscitelli, S., Kenney, R., Shulman, A., Medoff, D., Dixon, L.: Results from a pilot study of a computer-based role-playing game for young people with psychosis. Early Intervention in Psychiatry **13**(4), 767–772 (2019). https://doi.org/https://doi.org/10.1111/eip.12556

[63] Oppenauer-Meerskraut, C., Kropf, J., Bösendorfer, A., Gira, M., Heller, M., Lampel, K., Kumpf, A., Stamm, T.: I don't like gaming, it leads to a pressure to perform: Older adults refused gaming elements in a digital home-based physical training programme in two qualitative studies. In: CHIRA. pp. 178–184 (2017). https://doi.org/10.5220/0006516901780184

[64] Perry, Y., Werner-Seidler, A., Calear, A., Mackinnon, A., King, C., Scott, J., Merry, S., Fleming, T., Stasiak, K., Christensen, H., Batterham, P.J.: Preventing depression in final year secondary students: School-based randomized controlled trial. J Med Internet Res **19**(11), e369 (Nov 2017). https://doi.org/10.2196/jmir.8241

[65] Pesare, E., Roselli, T., Rossano, V.: Intelligent agent and virtual game to support education in e-health. In: DMS. pp. 204–211 (2015). https://doi.org/10.18293/DMS2015-35

[66] Podina, I.R., Fodor, L.A., Cosmoiu, A., Boian, R.: An evidence-based gamified mhealth intervention for overweight young adults with maladaptive eating habits: study protocol for a randomized controlled trial. Trials **18**(1), 1–14 (2017). https://doi.org/10.1186/s13063-017-2340-6

[67] Raminhos, C., Cláudio, A.P., Carmo, M.B., Gaspar, A., Carvalhosa, S., de Jesus Candeias, M.: A serious game-based solution to prevent bullying. International Journal of Pervasive Computing and Communications (2016). https://doi.org/10.1108/IJPCC-04-2016-0022

[68] Rego, P., Moreira, P.M., Reis, L.P.: Serious games for rehabilitation: A survey and a classification towards a taxonomy. In: 5th Iberian Conference on Information Systems and Technologies. pp. 1–6 (2010)

[69] Riaz, M.S., Cuenen, A., Janssens, D., Brijs, K., Wets, G.: Evaluation of a gamified e-learning platform to improve traffic safety among elementary school pupils in belgium. Personal and Ubiquitous Computing **23**(5), 931–941 (2019). https://doi.org/10.1007/s00779-019-01221-4

[70] Riedel, J.C., Feng, Y., Azadegan, A., Romero, M., Usart, M., Hauge, J.B.: Measuring the commercial outcomes of serious games in companies–a review. In: International conference on serious games development and applications. pp. 176–191. Springer (2014). https://doi.org/10.1007/978-3-319-11623-5_15

[71] del Río, N.G., González, C.S.G., González, R.M., Adelantado, V.N., Delgado, P.T., Fleitas, Y.B.: Gamified educational programme for childhood obesity. In: 2018 IEEE Global Engineering Education Conference (EDUCON). pp. 1962–1668 (2018). https://doi.org/10.1109/EDUCON.2018.8363476

[72] Sabben, G., Mudhune, V., Ondeng'e, K., Odero, I., Ndivo, R., Akelo, V., Winskell, K.: A smartphone game to prevent hiv among young africans (tumaini): Assessing intervention and study acceptability among adolescents and their parents in a randomized controlled trial.

JMIR Mhealth Uhealth **7**(5), e13049 (May 2019). https://doi.org/10.2196/13049

[73] Santos, P.A.: Deep gamification of a university course. In: Conference Proceedings SciTecIN. vol. 15 (2015)

[74] Sasupilli, M., Bokil, P., Wagle, P.: Designing a learning aid for dyslexic children. In: Research into Design for a Connected World, pp. 703–712. Springer (2019). https://doi.org/10.1007/978-981-13-5977-4_59

[75] Schell, J.: The Art of Game Design: A book of lenses. CRC press (2008)

[76] Seaborn, K., Fels, D.I.: Gamification in theory and action: A survey. International Journal of Human-Computer Studies **74**, 14–31 (2015). https://doi.org/https://doi.org/10.1016/j.ijhcs.2014.09.006

[77] Silva, J., Oliveira, E., Moreira, D., Nunes, F., Caic, M., Madureira, J., Pereira, E.: Design and evaluation of a fall prevention multiplayer game for senior care centres. In: International Conference on Entertainment Computing. pp. 103–114. Springer (2018). https://doi.org/10.1007/978-3-319-99426-0_9

[78] Smith, D., Ma, M., Jones, A., Unver, E.: None in three: The design and development of a low-cost violence prevention game for the caribbean region. In: Joint international conference on serious games. pp. 259–270. Springer (2017). https://doi.org/10.1007/978-3-319-70111-0_24

[79] Spook, J., Paulussen, T., Kok, G., van Empelen, P.: Evaluation of a serious self-regulation game intervention for overweight-related behaviors ("balance it"): A pilot study. J Med Internet Res **18**(9), e225 (Sep 2016). https://doi.org/10.2196/jmir.4964

[80] Stanmore, E.K., Mavroeidi, A., de Jong, L.D., Skelton, D.A., Sutton, C.J., Benedetto, V., Munford, L.A., Meekes, W., Bell, V., Todd, C.: The effectiveness and cost-effectiveness of strength and balance exergames to reduce falls risk for people aged 55 years and older in uk assisted living facilities: a multi-centre, cluster randomised controlled trial. BMC medicine **17**(1), 1–14 (2019). https://doi.org/10.1186/s12916-019-1278-9

[81] Stapinski, L.A., Reda, B., Newton, N.C., Lawler, S., Rodriguez, D., Chapman, C., Teesson, M.: Development and evaluation of 'pure rush': An online serious game for drug education. Drug and Alcohol Review **37**(S1), S420–S428 (2018). https://doi.org/https://doi.org/10.1111/dar.12611

[82] Timpel, P., Cesena, F.H.Y., da Silva Costa, C., Soldatelli, M.D., Gois Jr, E., Castrillon, E., Díaz, L.J.J., Repetto, G.M., Hagos, F., Castillo Yermenos, R.E., et al.: Efficacy of gamification-based smartphone application for weight loss in overweight and obese adolescents: study protocol for a phase ii randomized controlled trial. Therapeutic advances in endocrinology and metabolism **9**(6), 167–176 (2018). https://doi.org/10.1177/2042018818770938

[83] Vara, M.D., Baños, R.M., Rasal, P., Rodríguez, A., Rey, B., Wrzesien, M., Alcañiz, M.: A vr-based serious game to regulate joy in adolescents: A comparison of different devices. In: eHealth 360°, pp. 135–142. Springer (2017). https://doi.org/10.1007/978-3-319-49655-9_18

[84] Vieira, B., Pereira, L., Freitas, R., Terroso, M., Simoes, R.: A gamified application for assessment of balance and fall prevention. In: 2015 10th Iberian Conference on Information Systems and Technologies (CISTI). pp. 1–6 (2015). https://doi.org/10.1109/CISTI.2015.7170473

[85] Xu, F., Buhalis, D., Weber, J.: Serious games and the gamification of tourism. Tourism Man-

agement **60**, 244–256 (2017). https://doi.org/https://doi.org/10.1016/j.tourman.2016.11.020

[86] Zagal, J.P., Mateas, M., Fernández-Vara, C., Hochhalter, B., Lichti, N.: Towards an ontological language for game analysis. Worlds in play: International perspectives on digital games research **21**, 21 (2007)

[87] Zichermann, G., Cunningham, C.: Gamification by design: Implementing game mechanics in web and mobile apps. " O'Reilly Media, Inc." (2011)

# Part III

# Empower older adults to be more verbally resilient

# Introduction

## Motivation

In the previous part (Part II), I explored which vulnerabilities of older adults are already addressed by serious games and gamification. The review pointed out that most of the applications are about health vulnerabilities, mainly the risk for falling which is a risk caused by the physical frailty of elderly people. However, older adults do not only have vulnerable health, but their safety could also be at risk due to the increased vulnerability that comes with their age. Physical vulnerability strongly influences the fear of crime; a higher believed vulnerability results in a higher fear of crime [1]. Aspects of physical vulnerability that contribute to this were, among others, self-confidence and physical shape. In general, physical shape gets worse with age, thus the physical vulnerability increases with age.

A type of crime that older adults are specifically vulnerable to is doorstep scams. Doorstep scams are scams in which a con artist tells a fraudulent, yet believable, story, with the aim to gather personal information and/or belongings of the victim. Doorstep scams have a high impact on the victims, not only materially but also mentally and potentially physically. In this part, I explore how gaming technology can be used to make older adults more resilient against such scams. Doorstep scams can be seen as a form of social engineering: 'the exploitation of a human in order to gain unauthorized access to information' [4]. People can be manipulated with social engineering, as we generally tend to trust people. In doorstep scams this element of trust is also present, e.g. the con artist wears a (fake) uniform from a trusted company or organization. It is hard to prevent social engineering, however, spreading awareness among possible victims is an important way to make people less prone to falling for social engineering attacks [3]. To be resilient against doorstep scams it is important to know what to ask or say during a possible doorstep scam, but also the way these things are asked or said. Having an assertive tone of voice helps with being resilient against such scams. In this part, we refer to these elements together as verbal resilience.

## Approach

This part of the dissertation answers the subquestion:

> *How can interactive scenarios be applied in a serious game to make older adults more verbally resilient against doorstep scams?*

Virtual agents are autonomous entities in a virtual environment that look and behave like living organisms, such as human or animal characters [5]. A serious game with virtual agents can create a believable situation that mimics real-world doorstep scams to some extent. It is however also important to keep in mind the requirements of the target users. Therefore, a balance needs to be found between user-friendliness and accessibility, and believability of the scenarios. Together with the target group and domain experts in doorstep scam prevention, requirements are formulated, which are implemented in a prototype of a tablet application.

The goal of the resilience training is twofold. The first goal is that users learn about ways to react to a stranger at the front door, on the street or on the phone. To do so, I create six different interactive scenarios based on information gathered during field research. In these scenarios, users make decisions on how to respond to the con artist, which influences the progression of the scenario. To make it easier for users to apply the lessons learned in real-world situations, all scenarios follow a general outline based on doorstep scam scenarios encountered in the field research. Users receive feedback so that they learn to follow these steps. By repeatedly practising with this outline, it is expected that they will more likely use the learned steps in real life as well.

Next, users learn how to use their voice assertively. To do so, voice analysis software developed by Formolo [2] will be used. For this application, this software is trained to distinguish between assertive and non-assertive voices. While playing the scenarios, users choose one of the options during a choice moment and have to read this out loud as well. Their voice recording is then analysed and they immediately see the outcome of this analysis on the screen: the assertiveness of their voice is visually shown on an assertiveness meter that ranges from red (bad) to green (good). This is meant as a way to teach users to use their voice in a more assertive way, which is beneficial in preventing doorstep scams.

Adding voice analysis software offers a new way of controlling a scenario-based game, namely by the tone of your voice. Voice control is used in different situations, but the tone of voice as a way to progress in a serious game is a new way of control that is explored in this part of the dissertation.

Ultimately, the desired effect of the serious game is to make users more resilient against doorstep scams. However, quantifying and measuring this resilience is challenging. There is no test to access resilience, and it would be hard to design an experiment that is ethically sound to assess resilience in real-world scenarios. For this reason, I make the choice to evaluate the user experience and expected effectiveness using focus groups.

### Position according to Chapter 2

In this section, I position the work of this part with the use of the different terms defined in Chapter 2, see Figure 2.1 and Figure 2.2. The terms are written in italics.

For this project, the target group is an *age group*. Although everyone can become a victim of doorstep scams, older adults have a higher risk of becoming a victim and the impact is also higher for older adults. Their vulnerability can be characterized as a *safety risk*; doorstep scams can form a risk for both material damage, as well as physical or emotional damage. The empowerment method that is used is *training skills*. The skill that is trained is being resilient against doorstep scams, meaning that you are better able to defend yourself when you are approached by a con artist in a doorstep scam.

The serious game is specifically designed to be used on a *tablet* as this is the most suitable device for the target group, due to its screen size and adoption by older adults, and that suits the planned application. Different game mechanics are used in the serious game. To *reward* participants for their performance in a scenario, participants receive stars and points that reflect this. Moreover, they receive extensive *feedback*, tailored to their performance during the scenario. Players get *insight into their performance* by being able to see their previous scores and by the assertiveness meter of their voice analysis that is shown during each scenario. Finally, most *scenarios* include *avatars* to increase the believability of the scenario.

Intermediate versions of the serious game are evaluated in small *focus groups*. The final prototype is evaluated with *focus groups* of *19 participants* in total, that are domain experts or from the target group. Overall, the results were *positive*. Participants saw the added value of the application, mainly in larger prevention or information sessions.

**This part contains:**



# Bibliography

[1] De Donder, L., Verté, D., Messelis, E.: Fear of crime and elderly people: Key factors that determine fear of crime among elderly people in west flanders. Ageing International **30**(4), 363–376 (2005)

[2] Formolo, D., Bosse, T.: Towards interactive agents that infer emotions from voice and context information. EAI Endorsed Transactions on Creative Technologies **4**(10) (2017)

[3] Jain, A., Tailang, H., Goswami, H., Dutta, S., Singh Sankhla, M., Kumar, R.: Social engineering: Hacking a human being through technology. Journal of Computer Engineering (IOSR-JCE) **18**, 94–100 (10 2016). https://doi.org/10.9790/0661-18050594100

[4] Mouton, F., Leenen, L., Malan, M.M., Venter, H.S.: Towards an ontological model defining the social engineering domain. In: Kimppa, K., Whitehouse, D., Kuusela, T., Phahlamohlaka, J. (eds.) ICT and Society. pp. 266–279. Springer Berlin Heidelberg, Berlin, Heidelberg (2014)

[5] Vosinakis, S., Panayiotopoulos, T.: Simhuman: A platform for real-time virtual agents with planning capabilities. pp. 210–223. Intelligent Virtual Agents, Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-44812-8_17

# Chapter 3

# A serious game to improve the verbal resilience against doorstep scams

This chapter is published as: **van der Lubbe, L.M**, Gerritsen, C., & Bosse, T. (2020). A serious game to improve the verbal resilience against doorstep scams. International Journal of Serious Games, 7(2), 89-119.

**This chapter contains:**

## Acknowledgements

This project has been done in collaboration with the KBO-PCOB and has been funded by the Dutch ministry of Safety and Justice.

The development of the virtual training has been done with the help of Daniel Formolo and Marco Otte.

The original work described in Section 3.5.1, 3.5.2, and 3.5.3 is individually performed by students from the Vrije Universiteit Amsterdam; Celino Toussaint, Marvin Lau, and Dennis Walter. For this chapter, their work is summarized by the authors.

## Abstract

Doorstep scams, scams in which con artists tell convincing but fraudulent stories in order to enter the house of a victim and/or steal personal belongings or information, have a high impact on victims. These victims are often elderly people. Existing campaigns mostly focus on the behavioural aspects that help to prevent a doorstep scam from happening, but verbal skills also contribute to the resilience of possible victims. In this research, a serious game is designed and evaluated for this purpose. The results of the evaluation are positive. The serious game is specifically seen as a valuable addition to existing training meetings.

## 3.1 Introduction

Doorstep scams are scams in which a con artist has a convincing, but fraudulent, story with the purpose of coming into your house and/or stealing money or gaining personal information for fraudulent purposes. Doorstep scams frequently happen; numerous news reports about different stories exist. Elderly people are at a high risk of becoming victims of such scams. Since doorstep scams often have high emotional and economical impacts, various campaigns try to educate people on this topic in order to prevent doorstep scams from happening (e.g. ‘Spot it, Stop it’ [3] or ‘scam awareness month’ [6]). Such campaigns focus on behavioural aspects of the prevention of doorstep scams, such as opening a door with a chain on to prevent intrusion. However verbal skills (such as refusal assertiveness) are also important to be more resilient against doorstep scams. There is no known large scale prevention campaign that focuses on the conversation that takes place within a doorstep scam and the verbal skills needed to prevent these scams from happening. On a smaller scale, human actors are sometimes used to play a doorstep scam scenario with a group of people within, for example, larger group meetings about safety.

Doorstep scams are acknowledged by the Dutch Ministry of Safety and Justice as high impact crimes. Because of this, and because existing campaigns are insufficient, they funded research towards a virtual doorstep scam resilience training. The current research is executed in collaboration with a large Dutch organization for elderly called KBO-PCOB [5]. This research describes a virtual, simulation-based training, in the form of a serious game. This serious game can be used to improve the verbal resilience of the players, in order to minimalize doorstep scams from having negative outcomes. As the high-risk victims of such scams are elderly people, the intervention is aimed at this specific target group and therefore the design necessities of this specific target group are taken into account.

For the design of the system, theories from the domain of serious game research are used. We studied the hierarchy of players' needs [58]. This hierarchy, based on psychology, has different layers with needs of players that need to be successively fulfilled to increase the motivation of players. By using this hierarchy, we aimed to fulfil all those needs in order to increase the motivation of players. Moreover, literature about design principles for elderly users [59] is studied and used in the design of the app, to ensure that the app is accessible for the aimed target group. Such design principles are for example about recommended font size and contrast levels.

As mentioned, human actors can be used in small scale training sessions. An alternative to human actors can be virtual agents. Virtual agents are autonomous entities in a virtual environment that look and behave like living organisms (for example human or animal characters), that can interact with other human or virtual agent(s) in their environment [68]. Research in other domains (such as [11, 12]) has shown that the low costs and the degree of control of training with virtual agents are important advantages over training with human actors. Other advantages are for example that the training is available at any time and for a larger audience. When game elements are added to a virtual training, the virtual training becomes a serious game: 'a game in which education (in its various forms) is the primary goal, rather than entertainment' [41]. Previous research has shown that virtual trainings or serious games are effective learning resources [33, 46, 31, 66]. Moreover, virtual trainings can be repeated easily, and as many times as the trainee wants.

The objective of this chapter is to present the serious game that we have developed. This serious game aims to improve the players' verbal resilience against doorstep scams, by both teaching them more about what to say and how to use their voice in an assertive way. To achieve this goal, the serious game uses virtual agents in the role of scammer in interactive scenarios of doorstep scams and voice analysis to measure the assertiveness in the voice of a player. Moreover, this chapter presents a first evaluation of the system, with the goal to present insights in the perceived usability of the system by users from the target group. This evaluation gives insights in the interaction between the user and the system, as well as how the target group thinks that the system can be used after publishing. Moreover, the evaluation showed that using principles from theory in the design process of a serious game, is actually helpful. For example, the evaluation shows that all participants were capable of playing the serious game without assistance.

This chapter starts with a background section explaining more about the domain of doorstep scams and explores which virtual trainings and serious games already exist for skills that are relevant in the prevention of doorstep scams. Based on this research, Section 3.3 describes the design and implementation of a serious game to increase the verbal resilience against doorstep scams. In Section 3.4, a preliminary evaluation of this serious game and the results of the evaluation are described. In Section 3.5 some side projects that explore future work with (similar) serious games are described. Finally, Section 3.6 concludes the chapter with a summary of the system and some concluding remarks.

## 3.2 Background

To understand the domain of doorstep scams in the Netherlands, a field study was conducted to understand the content and progress of doorstep scams. For this field study the following sources were used: a focus group meeting with the partner KBO-PCOB, various conversations with domain experts, (news) articles and reports.

This section begins with a description of the findings of the field study in Section 3.2.1. Next, different skills, relevant for the resilience against doorstep scams, and existing (virtual) trainings and serious games for these skills are discussed in Section 3.2.2. Finally, the topic of serious gaming is introduced in Section 3.2.3.

### 3.2.1 Doorstep scam scenarios

Doorstep scams often happen at the front door, but can also happen on the street or via phone. During the field study, we identified three locations where doorstep scams often happen: at the front door, on the street, and on the phone. For each of these locations, we collected various frequently happening scenarios. An enumeration of these scenarios can be found in Appendix 3.A.

The fact that doorstep scams are a serious problem can be derived from the number of campaigns and news articles that can be found on the subject. However, it is hard to find statistics about the scope of the problem. This is due to the fact that there are multiple criminal activities associated with doorstep scams, which makes the registration inconclusive. Furthermore, often victims do not report a doorstep scam [40], due to shame for example.

### 3.2.2 Virtual training for verbal resilience

There is no previous work found in the domain of virtual trainings for doorstep scams. We therefore investigated which (virtual) training exist for skills that are relevant for doorstep scam resilience and which virtual training applications exist for the target group of elderly people.

#### 3.2.2.1 Assertiveness training

Being assertive means that you are behaving confident and that you dare to say what you think or believe [2]. This is a form of verbal resilience. Winship & Kelley [71] used a verbal response model to train assertiveness. Participants that were trained using this model showed an increase in their assertive behaviour. Research furthermore showed that assertiveness training within a group can also be effective [49]. It is also shown that verbal modelling and therapist coaching can increase the refusal of unreasonable requests [38, 39]. In other research [25] video feedback was found to be an effective approach for assertiveness group training.

**Saying no** Saying no, or refusal skills, are part of assertiveness trainings, applied in various domains among which are smoking behaviour, drugs usage, shoplifting and rape prevention. There are various ways in which somebody can say no: you can simply say no, make aversive statements, give a reason for not accepting an offer, change the subject or walk away [45]. For adolescents the most often used strategies are simply saying no to drugs [42, 22], alcohol [57, 52], or smoking [13, 65, 60], and giving a reason for not accepting the offer [57, 52]. In order to learn students to resist direct and indirect pressures to engage in negative behaviours, an effective prevention program should both show different verbal strategies as well as the need to be assertive when refusing an offer [42, 22, 65]. While nonverbal assertive skills can be used for different types of situations, practicing verbal strategies for specific situations is also needed [45].



Assertiveness is not only determined by the content of the message, but also by nonverbal aspects and the speech characteristics [21]. In this research, we focus on speech characteristics next to the content of the messages. Speech characteristics of assertiveness are for example speaking firmly or authoritative [44], duration of a reply [9, 19], medium latency of the response [21, 9], the loud volume of the voice [21, 9, 51], and the medium fluency [21, 51]. Another research points out that both the volume and the speed of the response are cues of confidence [32]. The Behavioural Assertiveness Test - Revised (BAT-R) [20] can assess assertive behaviour, via role play scenes. It studies different aspects in the behaviour of a participant: nonverbal aspects such as the number of smiles or the duration of eye contact, speech characteristics such as the characteristics mentioned before, and content aspects such as praise, appreciation and requests for new behaviour. However, the validation for this test is mixed [9]. One remark made is that for example the volume of voice is not unique for assertiveness. Within our research saying no, although often implicitly, is an important aspect of the assertive behaviour taught to the player. As suggested within the research of Nichols et al. [45] it is important to give verbal strategies for specific situations, this is done in this virtual training.

#### 3.2.2.2 Social skill training with virtual agents

Although some of the above-mentioned trainings are virtual, there are no examples mentioned yet of training programs for assertiveness with a virtual agent. However, virtual agents, sometimes in combination with virtual reality devices, are used for many different types of training programs, among which are social skills [10, 30]. An example is the *automated social skills trainer* (ASST) [61], a training in which human-agent interaction takes place via user speech and language interaction. The proposed training focuses mostly on communication, previous research has been done in the same field, however no comparable virtual trainings are found. *deLearyous* [66] is a serious game used to train interpersonal communication skills. Players learn how to use Leary's Rose in their advantage. The communication with the agent is in this case performed by unconstrained written language input. *Communicate!* [31] aims to train interpersonal communication skills. This game is used to train communication in a consultation setting. The interaction with the agent in this game is via multiple choice.

### 3.2.3 Serious games

When certain game elements are added to virtual trainings, they are considered as serious games. Goh et al. [28] give an overview of several strengths of serious games. One of these strengths is the fact that serious games offer covert learning, in combination with an already existing positive attitude towards games this makes serious games an easier accepted platform for learning. Furthermore, serious games are seen as fun and motivational since players want to achieve goals within the game. Another important strength of serious games is the feeling of control players have; they can practice a scenario as many times as they want to master a skill and they always have the option to shut down the tablet, giving them a feeling of safety. Greitzer [29] defines five levels of engagement for computer-based trainings (see Figure 3.1). These levels show different approaches to computer-based trainings, in which the level of engagement says something about the active role of the trainee and the complexity of the scenarios.

Siang et al. [58] created, based on the hierarchy of needs from Maslow [37], a hierarchy of players' needs (see Figure 3.2). This hierarchy is based on psychological theories and shows how psychology can be used in game design. The needs need to be addressed from the bottom to top successively. So, the rules need needs to be fulfilled before the safety need. With the use of this hierarchy, game designers can steer the motivation of players.

#### 3.2.3.1 Serious games for elderly users

Various types of serious games targeting elderly users exist to address different difficulties the target group faces. An often used type of game is the exergame, a game in which the player has to perform some sort of physical activity. Mostly these games are used to address problems that have to do with physical activity of the elderly users and related problems, such as balance and postural control. Often used technologies for such games are Nintendo Wii Fit sometimes together with the Balance Board (e.g. [54, 33]) or XBox Kinect (such as [54]). Exergames are also used for seniors with subsyndromal depression [62].

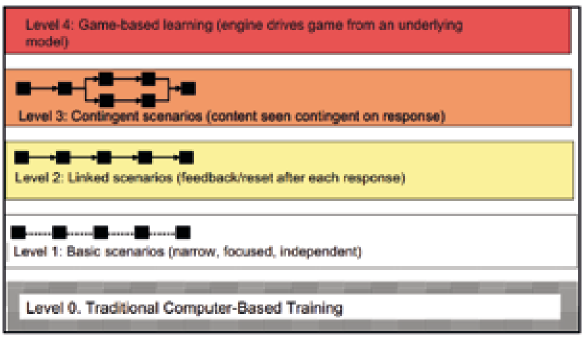



**Figure 3.1:** Levels of engagement for computer-based trainings [29]

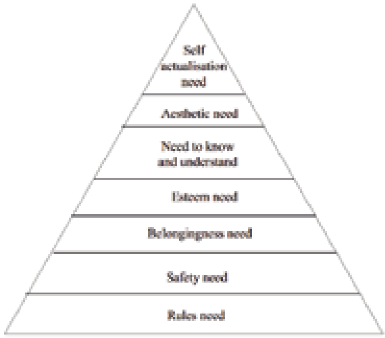



**Figure 3.2:** Hierarchy of players' need [58]

Serious games for elderly users can be used to address problems with cognitive abilities. So called brain training games, such as Brain Age [46], improve the elderly players' attention and memory skills. While these games are often designed especially to serve as a serious game, the research of Whitlock et al. [14] found that a non-serious game, World of Warcraft, improved the attention and spatial orientation of the elderly players as well.

Serious games for elderly do not only address physical or cognitive abilities, but it can also be used to enhance the users' social contacts, for example intergenerational interactions [56]. SilverGame [59] is a platform consisting of different activities to promote social activities, such as singing, dancing or performing fitness activities together with the use of video communication. The activities also serve as entertainment and promote exercise.

## 3.3 Design and implementation of serious game

This section describes the design and implementation of the virtual training that is developed during this research. An overview of the system is described in Section 3.3.1. In the next section, Section 3.3.2, the voice analysis module is explained. Section 3.3.3, describes the components of the training scenarios in more detail. First the scenarios are introduced (Section 3.3.3.1), followed by a section about the virtual agents and the environment (Section 3.3.3.2). Next, more about how the scores are calculated is described in Section 3.3.3.3. Finally, Section 3.3.4 describes the game layer that is built around the virtual training.

### 3.3.1 Overview of the system

The platform for which the virtual training is developed is a tablet. Since this is an often used platform by elderly people, KBO-PCOB posed this requirement.

Besides choosing a suitable platform for the target group, it is also important to keep the target group in mind when designing the game. Tailoring a serious game to the elderly target group reduces the risks of factors that reduce the efficiency of the game for this target group, such as information overload [33]. Not only the content and game play of the game must be tailored to the target group, also specific design principles [59] are used to adjust the game design to the target group. This includes high contrasts and a larger font type.

As explained earlier, serious games have some strengths, such as being fun and motivational. The virtual training will therefore be designed in the form of a serious game. Moreover, it was explained earlier that for assertiveness, the behaviour (way of using the voice) is also important. Hence, the application consists of three components: the voice analysis module, the training scenarios and the game layer.

The application starts with a main menu with four different options: scenarios, scores, explanation, and credits. On the scores pages the player can find their top ten scores and the average number of stars received for each scenario. It is also possible to reset all the scores.

For the application, six different scenarios were written, based on the field study (see Section 3.2.1). To ensure that the scenarios are credible, they have been evaluated with various domain experts provided by the KBO-PCOB, for example in a session with professional actors. Based on these evaluations, changes were made to the scenarios, which resulted in the six scenarios mentioned in Table 3.1.

Players can choose from these six scenarios, which can be played from a first person perspective. Moreover, the player can choose to play using voice analysis or not, in order to train the assertiveness of his/her voice. The voice analysis measures the assertiveness of the players' voice, and influences the scenario based on this measurement. The voice analysis is only available when playing with a network connection and is, if connected, by default turned on.

Figure 3.3 shows the flow of the application when playing a scenario. For each type of scenario (door, street or phone) a different intro is used. In the case of the door scenario the doorbell rings, the door is opened and the camera moves a bit forward. For the street scenarios the camera and an animated dog, placed close to the camera to represent the players' dog, move towards the avatar. In case of the phone scenarios a ring tone is played after which the screen of the phone placed in the environment changes, representing an incoming call. The end scenes of a scenario can either be closing the door (door scenarios), walking away (street scenarios), or a hung up sound (phone scenario).

**Table 3.1:** Scenarios used in the training

| | Location | Short description | Goal con artist |
|---|---|---|---|
| 1 | Front door | Energy meter check | Enter the house |
| 2 | Front door | Package delivery | Enter the house with package |
| 3 | Phone | Fraudulent back activity | Gain personal bank information |
| 4 | Phone | Lottery won | Gain personal information |
| 5 | Street | Money for charity | Collect money for non-existing charity |
| 6 | Street | Sell bracelet for charity | Sell overpriced/Fake bracelets |

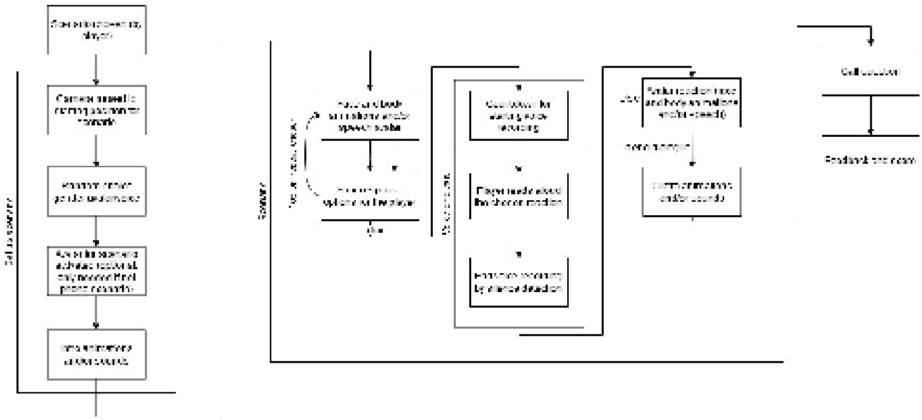


**Figure 3.3:** Flowchart of app flow

A scenario always starts with the virtual opponent. After the turn of the virtual opponent the player generally has four possible responses to choose from. One of these responses is to repeat the last turn of the agent (this response is only available if the virtual opponent has said something in the previous turn). This response is added to accommodate the target user group; in case they did not hear the agent correct. The other three responses influence the progress of the scenario.

When the scenario has ended the player is asked whether or not he/she would call the emergency number in such a situation. This question was added as during the field study it was found that victims often do not report a doorstep scam. Moreover, many people do not know that it is allowed to call the emergency number in case of a (suspected) doorstep scam. To motivate them to do so they are offered an opportunity to file a report in the application at the end of the scenario and they will receive feedback on this decision. After this the player will receive feedback on their choices during the scenario and see his/her score. For all feedback a ‘read aloud’ option is available. Section 3.3.3.3 discusses how the feedback and score are established.

### 3.3.2 Voice analysis module

As stated in Section 3.2.2 being assertive is not only represented in the content of your message, but also in the way the message is communicated. The application addresses both: the content is represented in the different choices provided by the interactive conversation, the voice analysis module addresses the influence of how the message is communicated. The technique used for this module is based on the research of Formolo and Bosse [23], the technical details of the module are beyond the scope of this chapter.

The voice analysis module is based on the Interpersonal Stances theory. This concept stems from social psychology, and can be defined as ‘the ways in which speakers and writers linguistically demonstrate their commitment to or attitudes about a person or proposition’ [26]. The module classifies the voice in 2 types of attitudes: Dominant (normally referred to as Above) and Submissive (Below). A modified version of the openSmile toolkit extracts the voice features, while an Support Vector Machine (SVM) algorithm classifies the extracted features into the categories. The SVM model was built using 4-fold cross validation over a dataset with 681 sentences of four people instructed in how to act into both categories. Details about the algorithm and the SVM tuning are described in [24]. The final accuracy of the module is 86.56%.

When voice analysis is performed, the recorded voice of the player is analysed by the algorithm. The application detects silence to determine when the recording ends. The ambient noise is measured while people are using the application. Silence is defined as a period of 3 seconds in which the volume is 20 decibels above the ambient noise volume maximum. The recording stops after 10 seconds or if silence is detected. The module runs on a server, it receives the recording of the user's voice and returns the classification status. In case of communication failure, the client application ignores the voice information and continues the dialogue without the voice analysis. The output of the module is the confidence percentage between the 2 categories (submissive or dominant), resulting in an assertiveness score.

The next section explains how the voice analysis is used within the training scenarios.

### 3.3.3 Training scenarios

The training scenarios consist of different components that together form interactive training scenarios with tailored feedback. These components are described in the next sections.

#### 3.3.3.1 Scenarios

The interactive conversations are the scripts of doorstep scams, represented in a conversation tree. Within these conversation trees, see an example in Figure 3.13 in Appendix 3.B, vertices represent either atomic agent behaviours or decision nodes (which enable the user to select a response), whereas edges represent transitions between nodes. They are turn-based, always starting and ending with the virtual agent. The blue rectangles show the avatar's dialogue. Each round the player is offered three choices as response to the agent (round rectangles in Figure 3.13). In general, these choices are of a good, moderate and bad level, influencing the conversation in the same way (represented in the colours of the round rectangles in Figure 3.13). Good in this context means assertive, bad means submissive. For some reactions of the player there are two outgoing arrows with conditions. These conditions are used by the voice analysis module and indicate what the influence of the assertiveness of the players' voice is on the progress of the scenario. In general, the outcomes of a conversation have the same mapping: good means no scam happened, and bad means that a scam happened. The red rectangle in Figure 3.13 shows a negative end state with the dialogue of the avatar.

When the module is turned on the player will not only be asked to make a choice within the scenario, the player is also asked to say this choice aloud and record this (within the application). The progress of the interactive scenario is then no longer only determined by the choices made, but also by the level of assertiveness measured in the voice of the

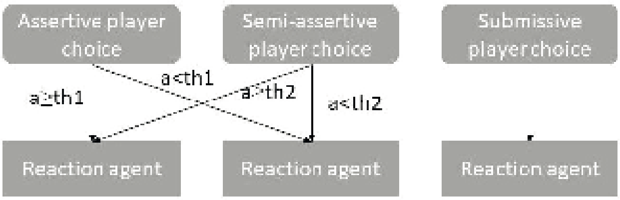



**Figure 3.4:** Flow scenario taking voice analysis into account



player. When the voice analysis module is not used, each reaction of the player directly leads to a reaction of the virtual agent. However, when the module is turned on in some cases one choice of the player can lead to two different reactions by the agent, determined by the level of assertiveness.

Figure 3.4 shows the flow of a choice moment when voice analysis is taken into account. When no voice analysis is performed the flow chart does not have thresholds and diagonal lines. The voice analysis module returns an assertiveness score (a). A high assertiveness score means a very assertive reaction. When a player has a high assertiveness score this will lead to the reaction normally given to an assertive player choice, although this might not be the most assertive choice when looking at the content. This also works the other way around. Threshold1 (th1) is lower than threshold2 (th2), since a very assertive player choice needs a little bit less assertive voice to have an assertive impact on the virtual agent.

Based on the information obtained during the field study a general outline is made for all the scenarios:

1. Background: getting to know the background of the story the con artist is telling.
2. Identity: getting to know about the identity of the con artist and his/her relation to the story that has been told.
3. Alternative: finding out alternatives to the suggested behaviour by the con artist.

In each scenario these three aspects can be found in this order. It depends on the specific scenario how these aspects are addressed. Following the same outline in each scenario gives the players guidance for other (real life) doorstep scams. Furthermore, it was discussed during the field research that it is important for people to get familiar with scenarios instead of only hearing about the theory of what to do to prevent/during a doorstep scam. It was mentioned that the fact that people have to read and speak certain sentences, such as ‘can I see your identification card?’, people might feel less of a burden to say this sentences in other situations as well. This supports our choice for scenarios that are somewhat similar in outline, so that these elements can be practised repeatedly and are easier to adopt for use in real world situations.

Figure 3.5 shows a screenshot of the application when the player has reached a choice moment, during the package delivery scenario at the door. At the bottom of the screen the four possible reactions are visible, the bottom one is the repeat option, the other three

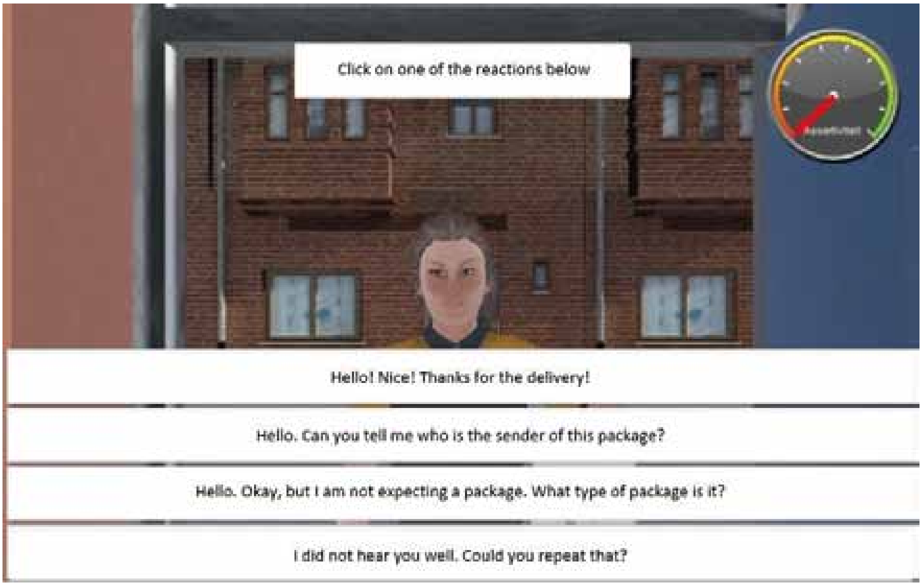



**Figure 3.5:** (Translated) screenshot of a choice moment

options are randomly ordered. In the upper right corner, the gauge for the voice analysis is showing the players last voice analysis score.

#### 3.3.3.2 Virtual avatars and environment

A three-dimensional environment is created using the game engine Unity [8], using various assets from the Asset Store [7]. Scripts for the functioning of the system are mostly programmed in C#. The environment features a small part of a residential area. Within this residential area there is a decorated street, that is used for the scenarios that take place on the street. Furthermore, one of the houses within the environment is partly furnished to feature the scenarios taking place at the front door and phone scenarios that take place in the living room. Figure 3.6 shows the different viewpoints within the environment. Players cannot move around in the virtual environment.

The virtual agents used within the scenarios, playing the role of doorstep scam artist, are modelled and animated using iClone [1]. For each scenario two similar agents are created, one female and one male. The outfits of the virtual agents are either recreated working uniforms, used to increase the credibility, or modern clothes. Figure 3.7 shows the eight avatars that are created for the door and street scenarios.

For the face animations, Facial Motion Capture [4] software, with a plug-in for iClone, is used. During two recording sessions a male and a female actor their face expressions were recorded while also recording the voice for the different dialogues. Their voices are also used for the scenarios that are not using a virtual agent (phone scenarios). After

**Figure 3.6:** Screenshots of different environments

recording, the facial animations needed some further editing because they were insufficiently realistic.

The flow of the application is programmed in a generic way, so that it is easy to add or change the content of the scenarios. The scenario specific content (avatars, dialogues, responses, tips, and feedback) are therefore stored in a database. By linking the code to the database the specific content is shown within the application. An overview of all the databases and the data stored in these databases can be found in Table 3.2.

**Figure 3.7:** Avatars created for door and street scenarios

#### 3.3.3.3 Feedback & scores

After each scenario the player will receive feedback. This feedback consists of a general reaction on the outcome achieved (whether or not you became a victim of a doorstep

**Table 3.2:** Overview of all used databases

| **Database** | **Description** |
|---|---|
| Avatar | This database contains the names of all the available avatars and links them to the right scene (location) and scenario number. |
| Dialogue | This database contains all the turns of the virtual opponent. Storing:<br>- Scene number, scenario number and the dialogue ID<br>- Text of the dialogue<br>- The tip variable IDs that is turned false in this dialogue (if applicable, more details on this in Section )<br>- The type of end state (if applicable)<br>- The body animation of the virtual avatar (if applicable) |
| Responses | This database contains the responses that are linked to the dialogues. Storing the following:<br>- Scene number, scenario number and the ID of the dialogue that the response is corresponding with<br>- Text of the response<br>- The ID of the default next dialogue<br>- The threshold for the speech analysis score (if applicable)<br>- The ID of the next dialogue when the speech analysis score is below the defined threshold, and the ID of the next dialogue when the score is above the threshold (if applicable)<br>- The action (animation) linked to the response (if applicable)<br>- The type of response (2=good, 1=average, 0=bad)<br>- A boolean if there is speech analysis for this response (default = true) |
| Tip | This database contains the tip variable IDs (corresponding with the dialogue table), their name and text. |
| Feedback | This database contains the different types of outcomes. The IDs correspond with the type of end state defined in the dialogue table. Furthermore, it contains the name and the feedback text of the outcomes. |

scam) and some general tips for the specific scenario. This is the feedback that is shown to the player by default. This feedback is stored in the feedback database (see Table 3.2).

Moreover, there is extended feedback. This feedback includes a paragraph that is specific for the scenario and is therefore independent of the outcome reached by the player. Furthermore, it includes feedback on the choice the player made for the emergency call question. There are three different feedback texts available for this. The first is for players who chose to call the emergency number, the second is for people who did not choose to call the emergency number but did have a negative outcome, and the last is for people who did not call the emergency number but did also not become a victim in the scenario.

The most tailored feedback are the tips. In each scenario a number of Boolean variables are defined, that are by default true. If a specific dialogue that is linked (which is stored in the dialogue database) to a variable is played during the scenario, this variable is set to false. For the variables that are still true at the end of the scenario a tip is given to the player. The tips are showed in such an order that they follow the progress of the scenario. An example is a variable about asking the identification, this is linked to the dialogue where the virtual opponent shows its identification. If this dialogue is not played during the scenario, the player receives feedback about this afterwards.

The tone of the feedback is positive. The goal of the feedback is to inform and educate players, not to punish them for wrong behaviour. This positive tone can be found in the general formulation of the sentences as well as in the fact that the feedback suggests other types of behaviour instead of telling the player what not to do. Since the general outline of the different scenarios is comparable, players can use the feedback for different scenarios. Furthermore, since the scenarios are comparable to real world doorstep scams, the feedback is also reusable in different possible real world doorstep scams.

Besides feedback the player also receives a score at the end of each scenario. The higher the score of the player, the better the performance during the training was. The highest score is 105. The score is calculated using the average score for the choices, the average score for the voice, a score for the result of the scenario, and a bonus (of 5 points) if the emergency number is called.

Next to a score a player also gets a number of stars (0-5) for a scenario. A player will receive one star if he/she earned more than 11 points during the game, two stars for more than 33 points, three stars for more than 55 points, four stars for more than 77 points, and the maximal number of five stars is achieved when more than 100 points are earned during the game.

### 3.3.4 Game layer

In Section 3.2.3 the levels of engagement of computer-based training programs were discussed. The described training aims at level three engagement, in which a limited amount of branching within the interactive scenario follows the choices of the player.

The game layer of the application is built around the actual training scenarios with the purpose of motivating players to (repeatedly) do the training. The scores and stars, discussed in Section 3.3.3.3 are one of the measures that are taken to do so. Moreover, a high score table in the app encourages players to improve their local scores. Siang et al. (2003) distinguished the different player needs. Those needs are addressed in the following way:

- Rules need: The players will receive an explanation about the rules of the game at the beginning, as well as instructions while they are playing.
- Safety need: The player will receive feedback at the end of each training scenario. The positive tone of the feedback will give them the feeling that they can handle the scenarios. Furthermore, when the player falls for a doorstep scam within the training this will not affect the player in real life.

- Belongingness need: The feedback that the players receive after each scenario helps them to improve their resilience against doorstep scams as they can use the feedback for other scenarios as well. The feedback will guide them to achieving better outcomes and achieve the goal of the game: to prevent a doorstep scam.

- Esteem need: The players have control over the scenarios since they can make choices influencing the progress of the scenario, as well as the option to quit a scenario. Furthermore, by the feedback received from the game the players will be encouraged, which will boost their esteem.

- Need to know and understand: By repeatedly training different scenarios players will improve their resilience skills, which gives them the ability to train with even more advanced scenarios that can be unlocked at a certain skill level.

- Aesthetic need: Several state of the art game development techniques have been used to build a 3D-environment with credible virtual agents.

- Self-actualisation need: (Advanced) players can play scenarios multiple times to test the reaction the virtual agent gives on different reactions, allowing them to test different reaction strategies.

## 3.4 Evaluation

Different consultations of domain experts have taken place during the development of the training. Once a first prototype was finished, this was evaluated during a focus group session. The results of this session are discussed in Section 3.4.1. Section 3.4.2 describes the setup and results of the final evaluation of the virtual training.

### 3.4.1 Focus group evaluation of prototype

During the design of the training, the partner organisation and its experts were repeatedly consulted. When a first prototype version was finished, a focus group with five elderly safety advisors of the KBO-PCOB (two females, three male) was organized to evaluate this version of the system. At the beginning of the focus group, a general instruction about the application was given. Next, participants could use the application in a private session for 30 minutes. After this, a group discussion was held to talk about the findings of the participants. This discussion was guided by open questions.

The overall reaction on the prototype was positive. Added value was especially seen in using the training scenarios in larger settings in which different users can interact with and help each other. Small remarks were made, that have resulted in some bugfixes and small textual changes for the final version of the application. Moreover, a bar was added to show players the remaining time for the voice recording.

### 3.4.2 Final evaluation

The final evaluation was in the form of two focus groups: one with eight elderly safety advisors and one with 11 elderly potential end users. This second group had a position as tablet coach for elderly for KBO-PCOB, which means that they are familiar with using a tablet. The elderly participants have ages between 68 and 84, with an average age of 74. Only 3 (16%) participants were female. Both sessions were organized on the same day, the 17th of July 2018, at the headquarters of the KBO-PCOB in Utrecht.



The focus groups started with a general explanation of the project, followed by 30 minutes of individual interaction with the system. After this, the participants provided us with qualitative feedback in a group discussion.

Next to this, participants filled in a questionnaire before the group discussion, which consisted of two parts. In the first part, some background information of the participants was asked, as well as some information about their experience with doorstep scams. The second part was based on the questionnaire used in a previous study of a virtual training [27]. The questionnaire consisted of 23 statements, about which the participants give their opinion using a 7-point Likert scale. The questions could be divided into four different categories:

- Content: five statements about the experienced realism of the scenarios and virtual avatars.
- Interaction: six statements about the believability of the interaction with the virtual avatars.
- Engagement: four statements about the engagement of the players with the scenario.
- Effect: eight statements about the opinion about the effect of the application.

All statements were identical to the ones used in [26], however three statements (Statement 14, 15, 16) were merged into one, as this was more suitable for this context. Moreover, five statements (Statements 10, 12, 16, 18, 22) were added to evaluate specific elements of the training, such as the voice analysis. A complete list of all the statements used within this evaluation can be found in Appendix 3.C.

Moreover, participants were given the opportunity to write down remarks after each question if they wanted to further elaborate on their rating further. Next to that there was some room for remarks at the end of the questionnaire.

#### 3.4.2.1 Results

Overall, the participants were positive about the application. In their opinion, the training was fun to do, and useful as well. Moreover, it was easy to use the application. The voice analysis clearly had added value for the participants. Again, some remarks were made to further improve the application. Minor changes, such as increasing font and button size and some minor textual changes, were made based on this before publishing the application.

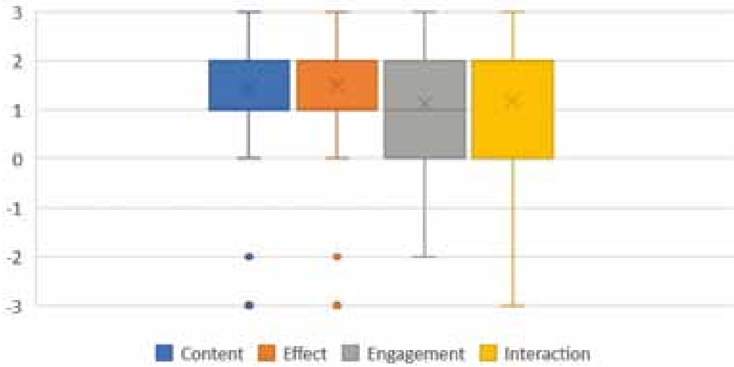



**Figure 3.8:** Boxplot for each topic in final evaluation

To analyse the results from the questionnaire, the answers were coded on a scale from -3 up until 3, numerically representing the answers on the 7-point Likert Scale that is used. If a question was not answered, these values are left blank.

Figure 3.8 shows the grouped results per category of the questionnaire. This figure shows that, in general, the participants (n=19) are average to strongly enthusiastic about the system. Mainly, people were positive about the added value of this medium in addition to the existing trainings. Moreover, the scenarios were found to be recognizable, and the interaction with the avatars was reasonable believable. The questions about engagement are rated a bit lower, but still above neutral. The highest average is found for the topic about effects, however the differences with the other topics are small.

The average score for each individual statement are also studied. This shows that all statements, but one (-0,25), are above neutral. This question was about the accuracy of the voice analysis. This question was only answered by seven participants, and of these participants two participants answered negatively to this question, one participant answered positively, and four answers were neutral.

The highest average score was given to the statement about the added value of the virtual training to role-plays with human actors (average score of 2,05).

Finally, the average scores for all statements in each topic are studied. For *content*, the averages can be found in Figure 3.9. This figure shows the previously discussed statement with the negative average. All other statements are evaluated with an average above one. The realism of the scenarios and the usefulness of the feedback are evaluated the highest. For the realism it was noted by a participant that in real doorstep scams, the scammers are more pertinacious, they want to hear arguments and to have a discussion if you counteract them.

For *effect*, the average scores are shown in Figure 3.10. One of the statements from this topic, is the statement that has the highest average of all statements, about the added value of the training to role-plays. This is the only average score of this topic as well as of the whole questionnaire that is evaluated with an average above 2. Another statement that is evaluated higher than other statements in this topic is about whether people liked

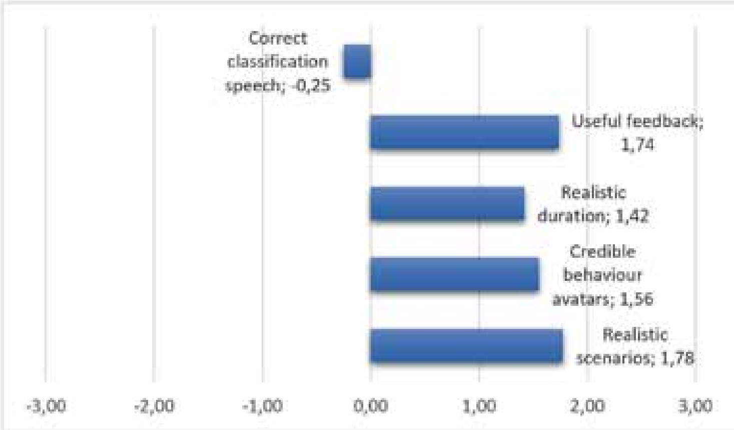



**Figure 3.9:** Average scores for content topic

to do the training. One of the participants noted as a remark to this question that he was surprised that it was such a learning moment for him. The lowest average score is for the statement about whether the voice analysis teaches participants about how they can use their voice in an assertive way.

For *engagement*, the average scores are shown in Figure 3.11. One statement has an average that differs more than 1 point from the other averages of this topic. This is a statement about whether participants felt personally addressed by the avatars. All other statements have an average above 1 and the highest average for the feeling of immersion. One of the participants noted that it felt like there was eye contact with the avatar.

Figure 3.8 showed that the average score for engagement was the lowest of the 4 categories. However, from Figure 3.11 it becomes clear that this low average is caused by the average score for the statement about whether participants felt personally addressed by the avatars. When this statement is left out of the average, it becomes much higher (1,44) which is the same as the average of the content topic.

Finally, for *interaction*, the averages are shown in Figure 3.12. All statements are evaluated with an average above 1, besides one statement. That statement is about whether wrong answers influenced the virtual avatars' behaviour. Another statement about the influence of users' choices on the behaviour of avatars was evaluated with a 1,11. It was noted that the virtual opponents could be more pressing. One participant reported to have noticed frustration after a resilient answer, the virtual avatar seemed frustrated. Whether or not a participant noted differences in the behaviour of the avatar might be influenced by the type of answers given. Whether the interaction with the avatars felt natural was rated with a 1,26. A participant suggested that using real actors would be better. The highest average in this topic is for whether there was always a suitable answer in the multiple choice menu. A similar average is given to whether participants felt that they could steer the course of the conversation.

In the group discussions during the final evaluation, the participants also noted that they


**Figure 3.10:** Average scores for effect topic

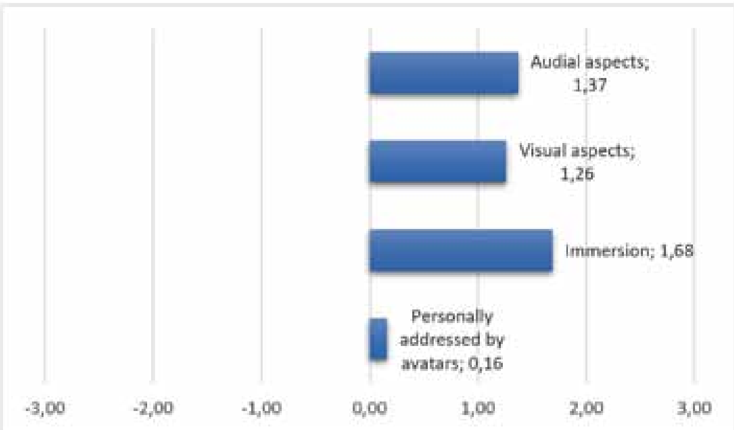



**Figure 3.11:** Average scores for engagement topic

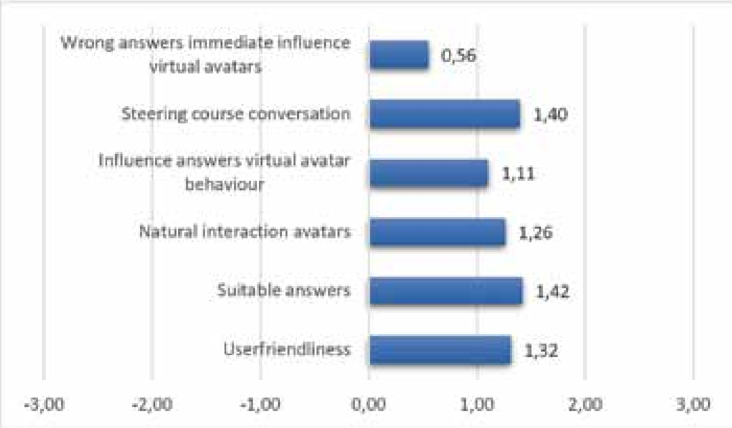



**Figure 3.12:** Average scores for interaction topic



see an added value of the app in prevention meetings with groups of elderly. Moreover, it was noted that the explanation of how the app works, as well as the answers, are clear for the user. Points for improvement that were noted had to do with making some components in the app clearer (for example fonts or button sizes). Moreover, some participants would like to see real people instead of avatars. Some remarks were made about the content of the scenarios, among which was a recommendation to use the name of the player in a scenario to make it more personal.

#### 3.4.2.2 Discussion of the results

The accuracy of the voice analysis module is not evaluated well, while during the training of the module the accuracy that was measured was high. This can be caused by different aspects. The performance can be lower in real-life situations compared to the situations of the voice samples used for training and testing the algorithm, for example due to background noise. Furthermore, we encountered some technical difficulties with the voice analysis, which might also have influenced the perception of the participants. Regarding the voice analysis, we learned that more testing and fine-tuning is needed to ensure a good accuracy in real-life settings. Moreover, from the evaluation it is clear that participants do not feel that they learned about an assertive use of voice. This might be improved by increasing the accuracy of the voice analysis as well.

Another point that was negatively evaluated was the statement about whether players felt personally addressed by the avatars. However, no additional feedback on this statement was given to explain this negative evaluation. Moreover, the other statements about engagement are evaluated higher, so it is unclear what causes this negative evaluation. One of the remarks that was made during the group discussion was that it might be more personal to use the name of the player in the scenarios. Another remark that has been made is that it would be better to use real actors instead of virtual characters. However, using actors would make the development more time consuming. In creating such serious games, the costs of development have to be considered in comparison to the believably and expected outcome. For future research, it is valuable to further study if the appearance of the virtual agents can be altered in such a way that the target groups finds them more realistic, and to study if this changes how players feel addressed.

## 3.5 Future work: extending the serious game with more game elements

The developed serious game is a fully functional, publically available tablet application. However, it is just a first version and it offers many opportunities for extension. As the game layer of this version only has limited game elements, different students worked on side projects to explore how other game elements could be integrated in this serious game as well[1]. This section describes an overview of their work, and the implications of their work for future extensions of the serious game, and in general for a virtual training in the form of a serious game.

### 3.5.1 Adding a quiz

In order to increase the learning value of the training as a whole, the first project explored the potential of adding a quiz to the training scenarios [64]. It is relevant to explore this, as the final evaluation of the serious game showed that the score for the statement about the learning value of the scenario was only 1.26.

The student compared his simplified version of a training scenario (control condition) with a version in which he added a quiz at the end (experimental condition). The quiz consisted of the following four components:

- True or false questions;
- Drag & drop questions in which three statements had to be placed in the right order;
- Matching questions in which a statement has to be matched with a scenario;
- Multiple choice questions.

In the evaluation of this project, 32 elderly participants took part. To evaluate the potential of adding a quiz, both objective and subjective measures were used. The objective measures concerned knowledge questions about the prevention of doorstep scams. Subjective measures concerned the perceived learning results, clarity, entertainment, motivation, challenge, and user-friendliness.

This work did not find an added value in the learning results of adding a quiz to a scenario game, but the research did indicate that the attitude of participants towards the version with the quiz was different. For perceived entertainment, motivation, and challenge, significant differences were found for the difference in average score between the two groups, in

[1] All students created their own simplified version of the virtual training to modify for their projects.

favour of the experimental condition. This indicates that adding an extra element might not directly contribute to better learning results, however it does not harm but in fact even improve the perceived entertainment, motivation, and challenge. This in turn could lead to more use of such a virtual training, which might increase the learning results on the long term. However, as the research only had a limited number of participants, that interacted with a simplified version for a limited amount of time, more research is needed to further explore this.



### 3.5.2 Adding different reward systems

Another approach to increase the learning outcome of the virtual training can be to add more reward systems to the serious game. However, while the project described in 3.5.1 was directly oriented on the knowledge (by adding a knowledge quiz), this student project aims to make the whole virtual training more gamified, and therefore more fun and motivation [35].

The student created three different designs for reward systems, based on literature research towards reward systems (for elderly) and learning and motivational theories [16, 18, 34, 36, 43, 50, 53, 55, 69], that could be added to the virtual training. These designs were discussed during one of the sessions at the KBO-PCOB in the form of a focus group (with six elderly participants that also took part in our focus group). The following three designs were evaluated:

1. A performance graph, showing the results from the last three attempts of the player. This is a reward system that aims for the individual user. In the performance graph the user sees a grade (0-10), showing the performance of the user in the scenario. Moreover, based on this grade, the player earns a trophy which can have three levels (bronze, silver or gold). While this gives feedback at the end of the scenario, another reward system within this prototype gives feedback during the scenario. That system is a bar that shows the level of openness/closeness of the player, indicating how close the player is to preventing or become the victim of a doorstep scam. The bar fluctuates as the player makes choices in the scenario.

2. The trophies from the first reward system are reused in this reward system. If a player earns a golden trophy, this gives the player an attempt in an edited version of hangman. The player can collect different words by playing games of hangman in between the scenarios. Next to this, in this version there are also goals to achieve for the player, which awards badges to the player.

3. In the third reward system a leaderboard is introduced, showing different values (high score, total score, number of prevented losses) to compare the users. During the scenario a bonus score is shown, that fluctuates with the choices made by the player.

During the focus group meeting, people were positive about the bar as it provides feedback during the scenario (prototype 1). However, remarks were made to increase the clarity of this mechanism. The trophies were also evaluated positively (prototype 1). The reactions to the performance graph (prototype 1) were more mixed, which was also the case for the hangman game (prototype 2). While participants liked it as it would be fun to play, they were worried that it might be too distracting. The badges (prototype 2) were seen as something fun, but the participants indicate to not have a desire to achieve them. Finally, the idea of a leaderboard (prototype 3) was liked, but the fluctuating bonus score during the scenario, introduced in prototype 3, was not.

Based on this feedback, the student decided to use the trophies and the hangman game, as well as the bar during the scenario, for the final prototype. Taking into account the remarks made during the focus group. In the control condition only the trophies were present, while in the experimental condition had all the reward systems of the final version implemented. In the final evaluation 54 elderly participants took part, 27 in each group. Objective data was collected, being the choices made in each scenario. Subjective data, both in the form of open and closed questions, was collected to give some background information to explain the objective data. No significant effects were found for an increase in the learning results for the experimental condition. In the subjective data it became clear that participants often did not pay attention to the bar during the scenarios and that people already had knowledge about doorstep scam prevention and might therefore not have learned enough to measure a difference. The results for the hangman component were mixed.

This study is an interesting example of how conventional game elements can be incorporated in new virtual training games as reward mechanism. Although the results do not show an added learning value, the idea of it is not rejected. Furthermore, feedback during the scenario was seen as a valuable addition to the serious game. Moreover, it shows that it is important to test if game elements are clear to users, as in the current design the bar missed its goal as it was overseen by many users.

### 3.5.3 Increasing motivation

While participants of the final evaluation of the serious game indicated that they like playing the game, the score for the statement about whether participants wanted to use the system in the future was lower. This project [70] aims to increase this score, by adding a reward mechanism to stimulate repeated play.

Similar to the research setup described in Section 3.5.2, three different prototypes of motivation mechanisms, based on literature research [15, 17, 18, 47, 48, 63, 67], were created and discussed during a focus gorup at the KBO-PCOB:

- A reward garden, shown in the menu of the game. New flowers to extend the garden appear when a scenario is completed, and is independent of the outcome of the scenario. This is used to stimulate repeated training.
- A score system in which the score is based on the choices made in the scenario.
- A dog mascot (cartoon style) was added to the scenarios to give immediate feedback, both in facial expression as well as textual, after a choice of the player.

The participants of the focus group were enthusiastic about the score system and especially enthusiastic about the dog mascot. They suggested adding a high score functionality to the game. While they were enthusiastic about the dog mascot, they were also worried that the immediate feedback would take the player out of the flow of the game.

The reward garden was received with less enthusiasm; it was unclear to the participants when new flowers appeared. Moreover, there would be a need to remove flowers at a certain moment, or have a maximum number of flowers as there is only a limited amount of space for these. The participants had mixed reactions on this, some were not enthusiastic about removing flowers, while others were neutral about the concept of gaining flowers.

For the final evaluation of this study, two versions of virtual training were created. In the control condition, a version was used that includes a score and high score system. For the experimental group this version was extended with the dog mascot. However, instead of giving immediate feedback, the dog mascot now only gives feedback at the beginning and end of a scenario, to address the concerns about the interruption of the flow of the app.

In the final evaluation, 62 participants, equally divided over the two groups, played the game and filled in a questionnaire afterwards. This questionnaire contained both questions about the background of the participants, experience with doorstep scams, and questions related to the experience with the game they played.

This research explored how motivational mechanisms around the scenario game can be used to increase the motivation of players to continue to play the game. Although it did not find any significant results, the focus group did point out that feedback during the game in the form of a mascot was received with enthusiasm. However, there were concerns about the interruption of the flow of the scenario due to this. The final evaluation showed that reducing it to only providing feedback at the beginning and end of a scenario did not stand out enough for players to notice it and to have an effect in this study. In the research described in Section 3.5.1, feedback during the scenario was also preferred by participants. Both studies also showed that to meet this desire a balance needs to be find between feedback and scenario flow. In future research, more extended attempts to do so can be researched.

### 3.5.4 Possibilities for future work

In the three described side projects, extensions for the virtual training are studied, focusing on the learning outcomes, rewards, and motivation. Although the studies only had a limited number of participants, the results point out some interesting things. Participants expressed a preference for feedback during the scenario when explicitly asked, but once implemented, this did not have an effect and was not even noticed by the players. However, when making additional components in the virtual training more outstanding, this might harm the flow of the app or make the interface (too) complex. This underlines the importance of (repeated) testing with possible end users to ensure that components are incorporated in the system in an optimal way, being noticeable without harming the flow of the app.

There are many other possibilities to extend the existing game. The game could be extended by increasing the number of scenarios or adding more game elements, such as social features. It could also be considered to replace the way to select one of the multiple choice reactions (touch) by voice control, which could simplify the flow of the system as no additional step is needed for the voice analysis. However, it needs to be studied whether this is an accessible way for elderly users to interact with the system.

It is also possible to use the key principles, the scenarios and the voice analysis, for a serious game on a different platform, such as a head-mounted display for a virtual reality (VR) version of the game. Currently, such technology is not commonly used among the aimed target group. Nevertheless, it is interesting to see if such a technology could increase the engagement of players. Moreover, it would be possible to also include a physical threat that comes with some types of doorstep scams. VR-technology allows for such things to be included, however the safety of the elderly target group needs to be guaranteed.

## 3.6 Conclusion

In this chapter the design, implementation and evaluation of a serious game to improve the verbal resilience of elderly users against doorstep scams is discussed. In collaboration with KBO-PCOB, different scenarios for this serious game were developed. The serious game aims to increase verbal resilience in two ways: via the content of the message and via the assertiveness of the voice of the user. For the design of the serious game, principles from theory were used. The hierarchy of players' needs was used to ensure that players are sufficiently motivated by the game. Moreover, design principles for an elderly target group are consulted. This made the game accessible for the target group, in the final evaluation all participants were able to use the application without assistance. During the final evaluation, 19 participants filled in a questionnaire about their experience with the virtual training. Overall, the evaluation of the system was positive, specifically the potential effect of the system was evaluated high, in particular the added value of the training to existing trainings/campaigns. Overall, the engagement category scored the lowest, but still positive.

The side projects that are introduced in Section 3.5 showed how different components, serving their own purpose, could be added to the virtual training to further enhance it. The research showed that there is a difficult balance between sticking close to the reality with realistic and uninterrupted scenarios versus giving feedback and helping players to make the right choices.

The long-term effect of the virtual training on people's performance is not studied in this research. This has different reasons, among which the fact that studying the effect of this doorstep prevention training is ethically challenging. It is always hard to measure potential situations, situations that might happen or not. To study the effect of this application, it is possible to study changes in the registered number of doorstep scams before and after the publication of the application. However, it remains difficult to draw conclusions on the effect of the application. There might be other reasons why the number decreased e.g. prevention by police, use of special locks on the door. It is, however, possible to test the knowledge of people before and after working with the application and again measure the effects after a longer period. Another possibility, that was beyond the scope of this project, is to test the reaction of participants in staged situations with the help of actors. However, the current study sheds light on a first important step that has been taken, it showed the feasibility of using a serious game for verbal resilience for the specific target group and it also gives more insight in how game elements can be used in this context and for this target group.

One of the important findings of this research is that participants saw a great added value of the training in incorporating it in existing training sessions. Moreover, this serious game shows how multiple choice interaction can be enriched with voice analysis. The application ('Trucs tegen babbeltrucs') is published and freely available in the Google Play Store and Apple's app store. Moreover, the application is used by the KBO-PCOB for prevention training sessions.



## 3.A Extended explanation scenarios

- Stories at the door:
  - The electricity, gas, or water, needs to be checked by the con artist, therefore he/she must enter your home.
  - The con artist has a delivery (package, flowers) that he/she wants to give to the victim. Either this is an excuse to enter the house or it is an excuse to seduce the victim to do a small payment (e.g. delivery costs), often with the purpose to steal more money from the victim than just the payment.
  - A con artist in the role of handyman wants to do some job for the victim, however he/she asks (a large amount of) money that needs to be paid right away. The job will not be done (correctly), but the money is taken.
- Stories on the phone:
  - The con artist calls with a story about fraudulent payments made, pretending to be the bank. He/she will ask private information, such as their debit/credit card PIN.
  - The con artist tells the victim that he/she has won a price, but in order to claim the price the victim needs to give information or make a small payment.
  - The con artist calls pretending to do a survey and asks the victim to give personal information. This information is later used to perform a financial scam or identity theft.
- Stories on the street:
  - The con artist claims to be collecting money or selling something for charity. Either the money is not used for charity or the con artist uses this to be able to easily pickpocket the target.
  - The con artist sells a newspaper or magazine, however the victim ends up with a long and expensive subscription.
  - The con artist sells some goods at the street, either the prize is too high (or the quality too low) or the good turns out to be completely worthless.

## 3.B Full flowchart (translated) example scenario

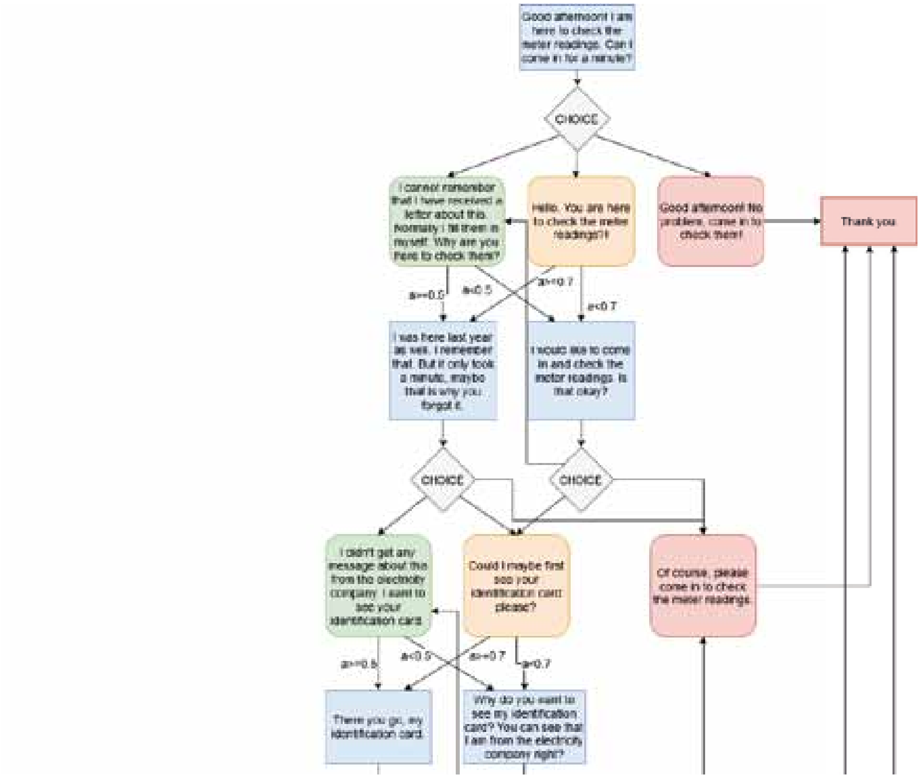



**Figure 3.13:** A (translated) scenario flowchart (part I)

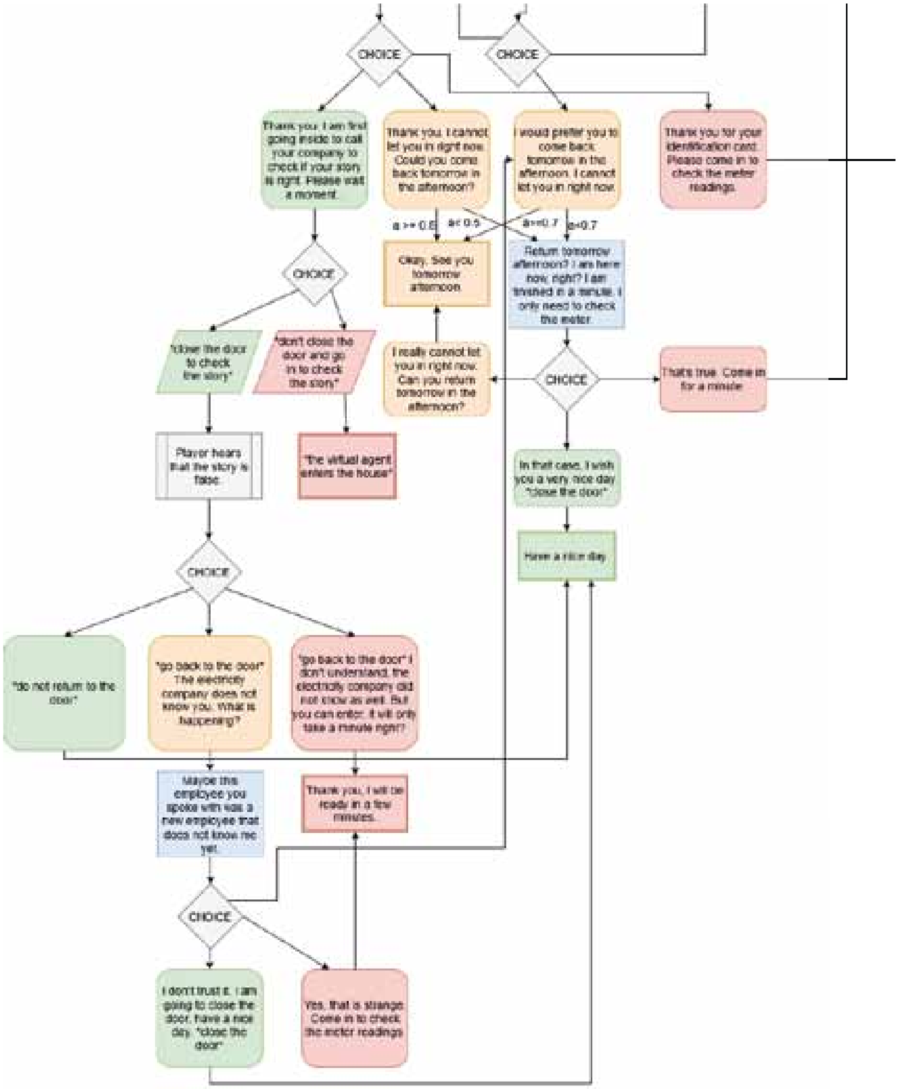



**Figure 3.13:** (part II) A (translated) scenario flowchart

## 3.C Questionnaire statements translated from Dutch

1. The trainings software was user friendly.
2. The multiple choices menus always contained an answer I agreed with.
3. I felt like I got better in handling the scenarios correctly.
4. I believe the events in the scenarios were realistic.
5. I believe the virtual characters acted credibly.
6. Interacting with the virtual characters felt natural.
7. I felt like my answers influenced the behaviour of the virtual characters.
8. I felt capable steering the course of the conversations.
9. When I said something wrong I directly noticed it from the behaviour of the virtual characters.
10. The duration of the scenarios was realistic.
11. I felt personally addressed by the virtual characters.
12. The feedback at the end of each scenario was useful.
13. During the training I felt immersed in the scenarios.
14. The visual aspects of the scenarios made me feel like the scenarios were real.
15. The audial aspects of the scenarios made me feel like the scenarios were real.
16. For people that tried the voice analysis only: I felt that the voice analysis classified my voice correctly.
17. This training made me think about the what to do and what not to do during a doorstep scam.
18. After following this training I am better capable to determine the right way of handling such situations.
19. For people that tried the voice analysis only: This training taught me how to use my voice in an assertive way during a doorstep scam.
20. I believe this training can be a useful addition to role-play training.
21. I liked doing the training.
22. The resilience training met my expectations.
23. I would like to continue doing this type of training in the future.

# Bibliography


[1] 3D Animations Software - iClone - Reallusion. www.reallusion.com/iclone/, accessed: 2020-01-30

[2] Assertive - meaning in the Cambridge Dictionary. www.dictionary.cambridge.org/dictionary/english/assertive, accessed: 2020-01-30

[3] Doorstep scams - Scams and fraud - Age UK. https://www.ageuk.org.uk/information-advice/money-legal/scams-fraud/doorstep-scams/, accessed: 2020-01-30

[4] Faceware Technologies, Inc. www.facewaretech.com, accessed: 2020-01-30

[5] Samen sterker voor senioren - KBO-PBOC. www.kbo-pcob.nl/, accessed: 2020-01-30

[6] Scams awareness month - Citizens Advice. www.citizensadvice.org.uk/about-us/campaigns/current_campaigns/scams-awareness-month/, accessed: 2020-01-30

[7] Unity Asset Store. https://assetstore.unity.com/, accessed: 2020-01-30

[8] Unity Real-Time Development Platform. www.unity3d.com, accessed: 2020-01-30

[9] Bellack, A.S., Hersen, M., Turner, S.M.: Role-play tests for assessing social skills: Are they valid? Behavior Therapy **9**(3), 448–461 (1978). https://doi.org/https://doi.org/10.1016/S0005-7894(78)80089-6

[10] Bosman, K., Bosse, T., Formolo, D.: Virtual agents for professional social skills training: An overview of the state-of-the-art. pp. 75–84. Intelligent Technologies for Interactive Entertainment, Springer International Publishing. https://doi.org/https://doi.org/10.1007/978-3-030-16447-8_8

[11] Bosse, T., Gerritsen, C.: Towards serious gaming for communication training - a pilot study with police academy students. pp. 13–22. Intelligent Technologies for Interactive Entertainment, Springer International Publishing. https://doi.org/10.1007/978-3-319-49616-0_2

[12] Bosse, T., Gerritsen, C., de Man, J.: An intelligent system for aggression de-escalation training. In: Proceedings of the Twenty-second European Conference on Artificial Intelligence. pp. 1805–1811. IOS Press (2016). https://doi.org/10.3233/978-1-61499-672-9-1805

[13] Charlton, A., Minagawa, K.E., While, D.: Saying "no" to cigarettes: a reappraisal of adolescent refusal skills. Journal of Adolescence **22**(5), 695–707 (1999). https://doi.org/https://doi.org/10.1006/jado.1999.0264

[14] Chen, S.T., Chiang, I.: Effects of improvement on selective attention: Developing appropriate somatosensory video game interventions for institutional-dwelling elderly with disabilities. Turkish Online Journal of Educational Technology-TOJET **11**(4), 409–417 (2012)

[15] Csikszentmihalyi, M., Khosla, S., Nakamura, J.: Flow at work. The Wiley Blackwell Handbook of the Psychology of Positivity and Strengths-Based Approaches at Work pp. 99–109 (2016). https://doi.org/10.1002/9781118977620.ch7

[16] De Schutter, B., Vanden Abeele, V.: Meaningful play in elderly life. In: Annual Meeting of the International Communication Association, Date: 2008/01/01-2008/01/01, Location: Montreal, Quebec, Canada

[17] Deary, I.J., Der, G.: Reaction time, age, and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples. Aging, Neuropsychology, and Cognition

**12**(2), 187–215. https://doi.org/10.1080/13825580590969235

[18] Deci, E., Ryan, R.: Intrinsic motivation and self-determination in human behavior: Springer science & business media (1985)

[19] Eisler, R.M., Hersen, M., Miller, P.M.: Effects of modeling on components of assertive behavior. Journal of Behavior Therapy and Experimental Psychiatry **4**(1), 1–6 (1973). https://doi.org/https://doi.org/10.1016/0005-7916(73)90030-X

[20] Eisler, R.M., Hersen, M., Miller, P.M., Blanchard, E.B.: Situational determinants of assertive behaviors. Journal of Consulting and Clinical Psychology **43**(3), 330–340 (1975). https://doi.org/10.1037/h0076717

[21] Eisler, R.M., Miller, P.M., Hersen, M.: Components of assertive behavior. Journal of Clinical Psychology **29**(3), 295–299 (1973). https://doi.org/10.1002/1097-4679(197307)29:3<295::aid-jclp2270290305>3.0.co;2-9

[22] Englander-Golden, P., Elconin, J., Satir, V.: Assertive/leveling communication and empathy in adolescent drug abuse prevention. Journal of Primary Prevention **6**(4), 231–243 (1986). https://doi.org/10.1007/bf01330264

[23] Formolo, D., Bosse, T.: Human vs. computer performance in voice-based recognition of interpersonal stance. pp. 672–686. Human-Computer Interaction. User Interface Design, Development and Multimodality, Springer International Publishing. https://doi.org/10.1007/978-3-319-58071-5_51

[24] Formolo, D., Bosse, T.: Towards interactive agents that infer emotions from voice and context information. EAI Endorsed Transactions on Creative Technologies **4**(10) (2017). https://doi.org/10.4108/eai.4-9-2017.153054

[25] Galassi, J.P., Galassi, M.D., Litz, M.C.: Assertive training in groups using video feedback. Journal of Counseling Psychology **21**(5), 390–394 (1974). https://doi.org/10.1037/h0037094

[26] Gales, T.: Identifying interpersonal stance in threatening discourse: An appraisal analysis. Discourse Studies **13**(1), 27–46 (2011). https://doi.org/10.1177/1461445610387735

[27] Gerritsen, C., Bosse, T.: Virtual communication training for the law enforcement domain. EAI Endorsed Transactions on Creative Technologies **5**(14) (2018). https://doi.org/10.4108/eai.16-1-2018.153642

[28] Goh, D.H., Ang, R.P., Tan, H.C.: Strategies for designing effective psychotherapeutic gaming interventions for children and adolescents. Computers in Human Behavior **24**(5), 2217–2235 (2008). https://doi.org/https://doi.org/10.1016/j.chb.2007.10.007

[29] Greitzer, F.L.: Ingredients of effective and engaging online learning. In: Inter-Lab 2005 Conference

[30] Howard, M.C., Gutworth, M.B.: A meta-analysis of virtual reality training programs for social skill development. Computers & Education **144**, 103707 (2020). https://doi.org/10.1016/j.compedu.2019.103707

[31] Jeuring, J., Grosfeld, F., Heeren, B., Hulsbergen, M., IJntema, R., Jonker, V., Mastenbroek, N., van der Smagt, M., Wijmans, F., Wolters, M., van Zeijts, H.: Communicate! - a serious game for communication skills -. pp. 513–517. Design for Teaching and Learning in a Networked World, Springer International Publishing. https://doi.org/10.1007/978-3-319-24258-3_49

[32] Kimble, C.E., Seidel, S.D.: Vocal signs of confidence. Journal of Nonverbal Behavior **15**(2), 99–105 (1991). https://doi.org/10.1007/bf00998265

[33] Konstantinidis, E.I., Billis, A.S., Mouzakidis, C.A., Zilidou, V.I., Antoniou, P.E., Bamidis, P.D.: Design, implementation, and wide pilot deployment of fitforall: An easy to use exergaming platform improving physical fitness and life quality of senior citizens. IEEE Journal of Biomedical and Health Informatics **20**(1), 189–200 (2016). https://doi.org/10.1109/JBHI.2014.2378814

[34] Landers, R.N., Bauer, K.N., Callan, R.C., Armstrong, M.B.: Psychological Theory and the Gamification of Learning, pp. 165–186. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-10208-5_9

[35] Lau, H.: The effect of reward systems in the gamification of doorstep scams simulation training for elderly (2018)

[36] Locke, E.A., Shaw, K.N., Saari, L.M., Latham, G.P.: Goal setting and task performance: 1969-1980. Psychological Bulletin **90**(1), 125–152 (1981). https://doi.org/10.1037/0033-2909.90.1.125

[37] Maslow, A.H.: "motivation and personality", 3 sub edition. New York (1987)

[38] McFall, R.M., Lillesand, D.B.: Behavior rehearsal with modeling and coaching in assertion training. Journal of Abnormal Psychology **77**(3), 313–323 (1971). https://doi.org/10.1037/h0031010

[39] McFall, R.M., Marston, A.R.: An experimental investigation of behavior rehearsal in assertive training. Journal of Abnormal Psychology **76**(2), 295–303 (1970). https://doi.org/10.1037/h0030112

[40] van der Meer, P.: Onderzoeksrapport veiligheid 2017 (version 1.0). Technical Report commissioned by KBO-PCOB (2017)

[41] Michael, D.R., Chen, S.L.: Serious Games: Games That Educate, Train, and Inform. Muska & Lipman/Premier-Trade (2005)

[42] Miller-Day, M.A., Alberts, J., Hecht, M.L., Trost, M.R., Krizek, R.L.: Adolescent relationships and drug use. Psychology Press (2014)

[43] Nagle, A., Wolf, P., Riener, R., Novak, D.: The use of player-centered positive reinforcement to schedule in-game rewards inreases enjoyment and performance in a serious game. International Journal of Serious Games **1**(4), 35–47 (2014). https://doi.org/10.17083/ijsg.v1i4.47

[44] Nichols, T.R., Birnel, S., Graber, J.A., Brooks-Gunn, J., Botvin, G.J.: Refusal skill ability: An examination of adolescent perceptions of effectiveness. The Journal of Primary Prevention **31**(3), 127–137 (2010). https://doi.org/10.1007/s10935-010-0213-9

[45] Nichols, T.R., Graber, J.A., Brooks-Gunn, J., Botvin, G.J.: Ways to say no: Refusal skill strategies among urban adolescents. American Journal of Health Behavior **30**(3), 227–236 (2006). https://doi.org/10.5993/AJHB.30.3.1

[46] Nouchi, R., Taki, Y., Takeuchi, H., Hashizume, H., Akitsuki, Y., Shigemune, Y., Sekiguchi, A., Kotozaki, Y., Tsukiura, T., Yomogida, Y., Kawashima, R.: Brain training game improves executive functions and processing speed in the elderly: A randomized controlled trial. PLOS ONE **7**(1), e29676 (2012). https://doi.org/10.1371/journal.pone.0029676

[47] Olds, J., Milner, P.: Positive reinforcement produced by electrical stimulation of septal area

and other regions of rat brain. Journal of Comparative and Physiological Psychology **47**(6), 419–427 (1954). https://doi.org/10.1037/h0058775

[48] Pairoa, I., Arunrangsiwed, P.: The effect of brand mascots on consumers' purchasing behaviors. World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering **10**(5), 1612–1615 (2016)

[49] Rathus, S.A.: An experimental investigation of assertive training in a group setting. Journal of Behavior Therapy and Experimental Psychiatry **3**(2), 81–86 (1972). https://doi.org/https://doi.org/10.1016/0005-7916(72)90003-1

[50] Richter, G., Raban, D.R., Rafaeli, S.: Studying Gamification: The Effect of Rewards and Incentives on Motivation, pp. 21–46. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-10208-5_2

[51] Ridgeway, C.L.: Nonverbal behavior, dominance, and the basis of status in task groups. American Sociological Review **52**(5), 683–694 (1987). https://doi.org/10.2307/2095603

[52] Rohrbach, L.A., Graham, J.W., Hansen, W.B., Flay, B.R., Johnson, C.A.: Evaluation of resistance skills training using multitrait-multimethod role play skill assessments. Health Education Research **2**(4), 401–407 (1987). https://doi.org/10.1093/her/2.4.401

[53] Sailer, M., Hense, J.U., Mayr, S.K., Mandl, H.: How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. Computers in Human Behavior **69**, 371–380 (2017). https://doi.org/https://doi.org/10.1016/j.chb.2016.12.033

[54] Schoene, D., Lord, S.R., Delbaere, K., Severino, C., Davies, T.A., Smith, S.T.: A randomized controlled pilot study of home-based step training in older people using videogame technology. PLOS ONE **8**(3) (2013). https://doi.org/10.1371/journal.pone.0057734

[55] Selcen, O., Sercan, Ş.: Gaining Reward vs. Avoiding Loss: When Does Gamification Stop Being Fun?, pp. 48–71. IGI Global, Hershey, PA, USA (2016). https://doi.org/10.4018/978-1-4666-8651-9.ch003

[56] Senger, J., Wälisch, T., John, M., Wang, H., Belbachir, A.N., Kohn, B., Smurawski, A., Lübben, R.Z., Jones, G.: Serious Gaming: Enhancing the Quality of Life among the Elderly through Play with the Multimedia Platform SilverGame, pp. 317–331. Springer Berlin Heidelberg, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-27491-6_23

[57] Shope, J.T., Copeland, L.A., Maharg, R., Dielman, T.E., Butchart, A.T.: Assessment of adolescent refusal skills in an alcohol misuse prevention study. Health Education Quarterly **20**(3), 373–390 (1993). https://doi.org/10.1177/109019819302000308

[58] Siang, A.C., Radha Krishna, R.: Theories of learning: a computer game perspective. In: Fifth International Symposium on Multimedia Software Engineering, 2003. Proceedings. pp. 239–245. https://doi.org/10.1109/MMSE.2003.1254447

[59] Srikwan, S., Jakobsson, M.: Using cartoons to teach internet security. Cryptologia **32**(2), 137–154 (2008). https://doi.org/10.1080/01611190701743724

[60] Sussman, S., Stacy, A.W., Dent, C.W., Burciaga, C., Burton, D., Flay, B.R.: Refusal assertion versus conversational skill role-play competence: Relevance to prevention of tobacco use. Statistics in Medicine **12**(3-4), 365–376 (1993). https://doi.org/10.1002/sim.4780120318

[61] Tanaka, H., Sakriani, S., Neubig, G., Toda, T., Negoro, H., Iwasaka, H., Nakamura, S.: Teaching social communication skills through human-agent interaction. ACM Trans. Interact. Intell. Syst. **6**(2), 1–26 (2016). https://doi.org/10.1145/2937757

[62] Taylor, L.M., Maddison, R., Pfaeffli, L.A., Rawstorn, J.C., Gant, N., Kerse, N.M.: Activity and energy expenditure in older people playing active video games. Archives of Physical Medicine and Rehabilitation **93**(12), 2281–2286 (2012). https://doi.org/https://doi.org/10.1016/j.apmr.2012.03.034

[63] Thompson, D., Baranowski, T., Buday, R., Baranowski, J., Thompson, V., Jago, R., Griffith, M.J.: Serious video games for health: How behavioral science guided the development of a serious video game. Simulation & Gaming **41**(4), 587–606 (2008). https://doi.org/10.1177/1046878108328087

[64] Toussaint, C.: Gamificatie: Het effect van het toevoegen van een quiz achter een serious game simulatie over babbeltrucs voor ouderen (2018)

[65] Turner, G.E., Burciaga, C., Sussman, S., Klein-Selski, E., Craig, S., Dent, C.W., Mason, H.R.C., Burton, D., Flay, B.: Which lesson components mediate refusal assertion skill improvement in school-based adolescent tobacco use prevention? International Journal of the Addictions **28**(8), 749–766 (1993). https://doi.org/10.3109/10826089309062171

[66] Vaassen, F., Wauters, J., Van Broeckhoven, F., Van Overveldt, M., Daelemans, W., Eneman, K.: delearyous: Training interpersonal communication skills using unconstrained text input. Proc. of ECGBL pp. 505–513 (2012)

[67] Van Calster, K., Lens, W., Nuttin, J.R.: Affective attitude toward the personal future: Impact on motivation in high school boys. The American Journal of Psychology **100**(1), 1–13 (1987). https://doi.org/10.2307/1422639

[68] Vosinakis, S., Panayiotopoulos, T.: Simhuman: A platform for real-time virtual agents with planning capabilities. pp. 210–223. Intelligent Virtual Agents, Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-44812-8_17

[69] Vroom, V.H.: Work and motivation. Work and motivation., Wiley, Oxford, England (1964)

[70] Walter, D.: Increasing motivation in elderly to continue playing serious games: A case study of doorstep scams prevention (2018)

[71] Winship, B.J., Kelley, J.D.: A verbal response model of assertiveness. Journal of Counseling Psychology **23**(3), 215–220 (1976). https://doi.org/10.1037/0022-0167.23.3.215

# Part IV

# Empower older adults for healthy ageing through nutrition

# Introduction

## Motivation

Older adults are at risk of being malnourished. This is a vulnerability for around 21% of the older adults that are living at home [1]. Malnourishment has various health risks and decreases the quality of life [2]. The PROMISS-project [1] aims to prevent malnutrition among older adults and to formulate recommendations for a healthy and active lifestyle. In particular, the focus is on the protein intake of older adults. Protein intake is related to frailty and physical functioning [4, 3]. Since many older adults live at home (90-95%), good physical functioning is an important attribute for their independence. Moreover, it supports a healthy and active lifestyle. The research that is described in this part of the dissertation is performed within the PROMISS-project [1]. More specifically, it was part of the Dutch PROMISS trial, a diet trial that was performed as part of the PROMISS-project.

Adhering to a healthy lifestyle or diet can be stimulated using smartphone or tablet applications. Such apps can for example be used to change eating or physical activity habits. When designing such apps, is important that the requirements of specific subgroups, such as older adult users, are taken into account [7]. Existing healthy eating apps that were evaluated by experts showed different usability problems for older adult users [8]. In general, it was found that technology that is specifically designed for an older target group can influence the initial evaluation and long term adoption of users [5].

## Approach

Within this part of the dissertation, I aim to address the subquestion:

> *How to design a gamified persuasive diet tracking system as part of a diet intervention for older adults?*

The PROMISS trial aims at increasing the protein intake of older adults with a relatively low protein intake [6]. The dietary advice studied in this trial aims to increase participants' protein intake. Products are valued according to a protein point system designed for this trial. The older adults participating in the trial are provided with a personalized diet plan and information on products and proteins. The diet tracking system that is designed for this trial includes the same information, yet on top of that, it offers users the possibility to track their dietary intake.

As a first step, an initial prototype is designed and used in a pilot study of the PROMISS-project. Results of this pilot are used to refine and extend the system requirements, together with new input from the involved dieticians about the final setup of the dietary advice. These new requirements are used to develop a new, more advanced, version of the application that would meet the user expectations and dietary guidelines in a better way. The design and evaluations of the pilot system are described in Chapters 4 and 5. Moreover, as described in Chapter 5, gamification is added to the application to stimulate the adherence and compliance of participants. When choosing the game mechanics, I had to take into account both the special needs of the participants, as well as the diet guidelines and the desired diet behaviour. Because the main focus of the PROMISS trial is on the protein-rich diet, it was important that the gamification should not interfere with the diet choices of the participants and should stimulate the same behaviour that is stimulated by the diet guidelines and have no side effects on other types of behaviour. Thus, it is important to consider the possible effects of each game element that is added to the application on the diet behaviour of the participants. Focus groups with participants from the pilot study are held to discuss the changes that are made to the prototype from the pilot before the application is used within the larger evaluation study.

Whether it is beneficial to use the system is assessed in Chapter 6. This is done by looking at the effects of using a diet tracking system for a period of a maximum of six months as part of the PROMISS diet trial. Benefits for the protein intake, the experience of the diet, and knowledge on protein are analysed to assess the overall feasibility and effectiveness of the final system.

The evaluated application is only slightly adaptive to the user's eating habits. Suggested meals are adapted once a certain adaptation is made multiple times by the user, for example, a product is added or replaced with another product. However, with the collected data it is possible to study how well AI techniques can help to make this adaptation process smarter to increase the user-friendliness of the application. This is studied in Chapter 7, where an intelligent workflow of algorithms customizes the meal recommendations of the system to the eating behaviour of users. Four different algorithms are used to create personalised suggested meals by using the diet preference (vegetarian/pescatarian/none) of a user, using popular product category combinations, using popular product item combinations, and using and dividing the remaining protein point at the end of the day. An association rule learning algorithm is then used to help find additional products when a user edits a suggested meal. The goal of this intelligent workflow is to reduce the time that users need to complete their meal registration. The time spent on this task is measured with the number of clicks needed to register a meal. With an evaluation based on historical data, collected in the six months evaluation study, it is studied to what extent such an intelligent workflow reduces the number of clicks.

### Position according to Chapter 2

In this section, I position the work of this part with the use of the different terms defined in Chapter 2, see Figure 2.1 and Figure 2.2. The terms are written in italics.

For this project, the target group is an *age group* as the study specifically targets older adults. A low protein intake has various health consequences; thus the targeted vulnerability is a *health risk*. The empowerment method is *supporting behaviour*: participants are supported in their daily diet behaviour.

The application is designed for *tablets*. Different game characteristics are used. *Objectives* are used in the form of protein goals and game point thresholds. *Rewards* are given to participants with achievements, the ability to play a mini-game, and game points. Moreover, *notifications* are sent to motivate participants to use the system and stick to their diet. Finally, the *mini-games* that serve as a reward are also used to increase the protein knowledge of users.

Intermediate versions of the system were evaluated in a *pilot study* and individual focus groups. The final application is evaluated in a larger study with an *experimental setup*.

For this study, *48 participants* used the system, divided over two conditions: the normal condition where the application is not enriched with gamification, and the gamification condition that extends the standard application with gamification elements. Overall, the experienced effectiveness, user-friendliness of the system, notification messages, and gamification are evaluated *positively*. Although *no effect on protein intake* was found, we did find a *high adherence* for using the tablet application. The results on the *effect of gamification on the knowledge* about protein points are inconclusive.

### This part contains:



### Acknowledgements

This work was supported by the European Union Horizon2020 PROMISS Project 'PRevention Of Malnutrition In Senior Subjects' (grant agreement no. 678732).

For this part a special thanks goes to the PROMISS trial dietitians: Merel Vrijmoeth, Mariska Bout, and Marjon Veeke.

## Bibliography

[1] Promiss. `http://www.promiss-vu.eu/`, accessed: 2022-01-26

[2] Hickson, M.: Malnutrition and ageing. Postgrad. Medical Journal **82**(963) (2006)

[3] Isanejad, M., Mursu, J., Sirola, J., Kröger, H., Rikkonen, T., Tuppurainen, M., Erkkilä, A.T.: Dietary protein intake is associated with better physical function and muscle strength among elderly women. British Journal of Nutrition **115**(7), 1281–1291 (2016)

[4] Kobayashi, S., Asakura, K., Suga, H., Sasaki, S.: High protein intake is associated with low prevalence of frailty among old japanese women: a multicenter cross-sectional study. Nutrition journal **12**(1), 164 (2013)

[5] Mitzner, T.L., Savla, J., Boot, W.R., Sharit, J., Charness, N., Czaja, S.J., Rogers, W.A.: Technology Adoption by Older Adults: Findings From the PRISM Trial. The Gerontologist **59**(1), 34–44 (09 2018). https://doi.org/10.1093/geront/gny113

[6] Reinders, I., Wijnhoven, H.A.H., Jyväkorpi, S.K., Suominen, M.H., Niskanen, R., Bosmans, J.E., Brouwer, I.A., Fluitman, K.S., Klein, M.C.A., Kuijper, L.D., van der Lubbe, L.M., Olthof,

M.R., Pitkälä, K.H., Vijlbrief, R., Visser, M.: Effectiveness and cost-effectiveness of personalised dietary advice aiming at increasing protein intake on physical functioning in community-dwelling older adults with lower habitual protein intake. BMJ Open **10**(11) (2020)

[7] Wang, Q., Egelandsdal, B., Amdam, G.V., Almli, V.L., Oostindjer, M.: Diet and physical activity apps: Perceived effectiveness by app users. JMIR mHealth uHealth **4**(2), e33 (Apr 2016). https://doi.org/10.2196/mhealth.5114

[8] Watkins, I., Kules, B., Yuan, X., Xie, B.: Heuristic evaluation of healthy eating apps for older adults. Journal of Consumer Health on the Internet **18**(2), 105–127 (2014). https://doi.org/10.1080/15398285.2014.902267

Chapter 4

# Designing a system with persuasive communication to improve diet compliance for elderly users

This chapter is published as: **van der Lubbe, L.M.**, & Klein, M.C.A. (2019). Designing a system with persuasive communication to improve diet compliance for elderly users. In Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare (pp. 234-241).

**This chapter contains:**

## Abstract

This chapter introduces a system with persuasive communication to improve the diet compliance and adherence of participants of the PROMISS-project. For this persuasive communication different strategies and ways of personalisation are used, in accordance with existing literature. Furthermore, as the target group is elderly users, the design is tailored to their specific needs. A first prototype was created, based on seven functional requirements. During a pilot study, the prototype is evaluated with seven participants. Based on lessons learned during this pilot, as well as new requirements from the PROMISS dietitians, a refined design of the system was implemented. This refined design is briefly evaluated with four participants of the pilot during individual interviews. The changes made to the prototype were evaluated positively.

## 4.1 Introduction

In Europe, 90-95% of the elderly people are living at home. Of these elderly, 21% is malnourished or at risk of being malnourished [1]. While ageing, the body composition changes, which increases the risk of malnutrition. However, there are many other factors contributing to the malnutrition of elderly people, resulting in various health risks and a decreased quality of life [13]. The PROMISS-project [1] aims to prevent malnutrition among elderly and to formulate recommendations for a healthy and active lifestyle. More specifically, they research a diet to improve the protein intake of elderly. A low protein intake is associated with frailty, as found in a research in Japan among elderly women [17]. Another research [15] supports this; it found a positive influence of protein intake on the physical functioning of elderly women. Physical functioning is important to maintain an active and healthy lifestyle for elderly, which is important since the majority of the elderly people stay at home. Research towards effective interventions, for which different forms are possible, is therefore useful.

In previous research it is shown that, comparing a smartphone, a website and a paper diary intervention, the adherence was the highest for the participants using the smartphone intervention [4]. Furthermore, other researches showed that a smartphone app improved the diet compliance of participants, e.g. [7].

In order to meet the specific needs of a target group, it is advisable to create a food journal design specific for this group, rather than a generic journal [6]. Therefore this chapter describes the design and initial evaluation of a system with persuasive communication in order to promote diet compliance and adherence, to improve the protein intake of the elderly target group of the PROMISS-project.

Background literature about persuasive communication and application design for elderly users is discussed in Section 4.2. In order to develop the system a prototyping approach will be used. Section 4.3 discusses the prototype and its evaluation. A refined system is developed (Section 4.4) and re-evaluated with four participants (Section 4.5). Finally, in Section 4.6 discusses the current research and some possible extensions.

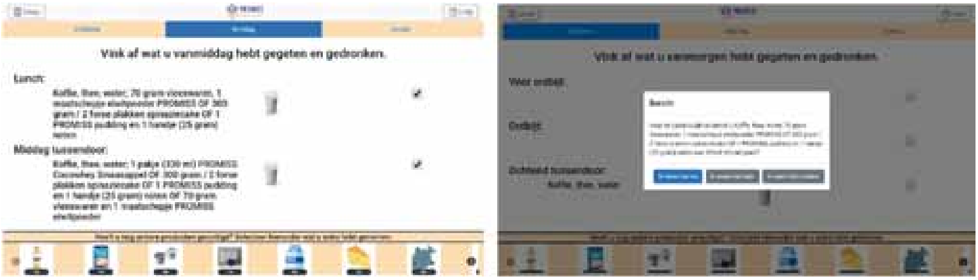



**Figure 4.1:** Screenshots of the initial version of the application



## 4.2 Related literature

As the goal of the system is to improve diet compliance and adherence by the means of persuasive communication, Section 4.2.1 describes related literature about persuasive communication in the health domain. Furthermore, as the target group of the application is elderly people, Section 4.2.2 will explain what aspects need to be taken into account when designing for this target group.

### 4.2.1 Persuasive communication

Persuasive technology aims to change the user's attitude or behaviour [14]. Computers have multiple advantages over human persuaders, among which are their anonymity, persistency, and the access to areas where human persuaders cannot come or are not welcomed to [8]. Furthermore, as shown in the introduction, smartphone applications are an effective means to increase participants' diet compliance. Being compliant to a diet often requires some form of behaviour change, persuasive communication is therefore useful in such an application. To study their compliance and behavior, users need to log their behavior in a system. In order to increase how often users log something in a system, notifications have found to be an effective method [3].

There are different ways in which messages can be tailored to improve their persuasiveness: framing, politeness strategies, and persuasion principles. Framing can be done in two ways: gain framing has the emphasis on the benefits, while loss framing has the emphasis on the costs. In a research towards the attention and response efficacy of participants for a prevention message, it was found that tailoring the type of framing used with the gender of the participant gives better results [5]. Gain framing was found to be more effective for female participants, whereas male participants responded more on messages with loss framing.

Messages should not only be persuasive, but also polite. In the research of [11] the politeness and persuasiveness of different politeness strategies is researched among an elderly target group. Among the tested strategies were *direct command*, *indirect suggestion*, and *questions*. *Questions* has the highest combined score, and can therefore be seen as the most polite and persuasive strategy.

Persuasive messages can be written taking into account Cialdini's six principles of persuasion [18]: *reciprocation, commitments and consistency, consensus or social proof, liking, authority*, and *scarcity*. The personality of a person can be described by using the Five-Factor Model [9], in which a personality is described by five traits: *agreeableness*, *conscientiousness*, *emotional stability*, *openness*, and *extraversion*. It has been researched what the influence of personality is on the effectiveness of healthy eating messages constructed using the Cialdini's principles using both positive and negative framing [19]. It is found that the *authority* principle is the most effective for most traits. *Consensus* messages are better for people with a low score on *conscientiousness*, while *commitment* messages work better for people with a high *conscientiousness*. Furthermore, people with a low *openness* prefer the *consensus* principle over *commitment*. In general, positive framing scored better compared to negative framing, but this was especially the case for people with a high *conscientiousness* or high *emotional stability*.

### 4.2.2 Application design for elderly users

In order to make a design that is suitable for the elderly target group, existing guidelines are researched. Tailoring the design to the elderly user group has found to be good for, among other things: the effectiveness, efficiency and user satisfaction [16]. This might however increase the task times for younger or more able people. Since there is a wide variety of elderly users and their experience with technology, it is important to find a balance when designing the application.

For the look and feel of a touch-based application it is important to have larger screens, with more space between buttons, and larger fonts [2]. Some other important things noted are using a separate keyboard and numpad in the application, using labelled icons, and using buttons instead of a menu structure.

When looking at the ways that elderly prefer to interact with touch-based applications it is found that they prefer tabbing over drag and drop actions [2]. Furthermore, it is important to have clear navigation in the application, which includes showing users where they are in the application and always giving them clear ways to progress, close, or go back through the interaction [12]. Because of the limited screen size of mobile devices, there is limited space for the needed information. For elderly it is important to show only relevant information on a page and also reduce scrolling to a minimum [12].

On a more general level, for no specific target audience, it was found in a qualitative research on using a smartphone application for diet and nutrition that users prefer quick and easy administration of their food intake, and the users want to be made aware of their intake and weight management [7].

## 4.3 Prototyping phase

The development of the system starts with the prototyping phase. In order to get effective feedback, a first prototype of the system is created and evaluated during one of the pilots within the PROMISS-project. The feedback that is gathered during this pilot is than used

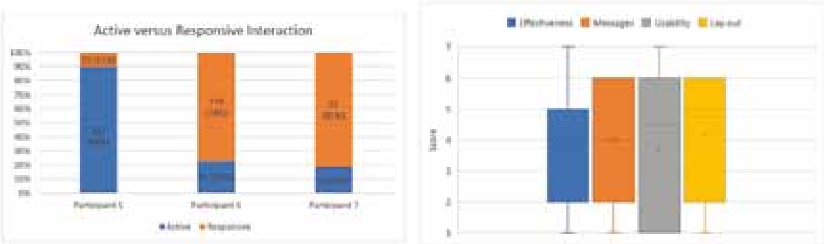



**Figure 4.2:** Rates of active / responsive interactions (left) and boxplot of questionnaires outcomes of six participants (right)

in order to improve the prototype, which results in a refined system discussed in Section 4.4.

The prototype of the persuasive communication system is build as a tablet application. Based on the intervention designed in the PROMISS-project, the functional requirements (req.) of the prototype are the following:

1. The system should give an overview of the personalised diet plan, per day. The diet plans are provided by the dietitians of the PROMISS-project. This diet plan is the same for each day.
2. The user receives notifications as a reminder to take a meal from the diet.
3. The user can enable a reminder for a notification that he/she has received.
4. The user can enter whether he/she has followed an advice.
5. The user can add extra items that he/she has consumed outside of the diet.
6. The lay-out and user interactions are suitable for elderly users.
7. The notifications are tailored to the personal characteristics of the user.

Figure 4.1 shows the prototype application. On the left is an example of the diet plan for a participant in the afternoon. The screenshot on the right shows a notification that is send to the user as a reminder for a specific meal.

### 4.3.1 Pilot study

The main goal of the pilot is to test the functioning of the application (e.g. find bugs and difficulties), as well as the interaction between the users and the system, and the user experience. In order to gain information about these aspects the participants received a Lenovo Tab4 10 (with an Android operating system) with the application, for the duration of three weeks.

With the use of logfiles from the application it is possible to study the user profiles by looking at the responsive and active interactions of users with the system. Due to time limits, the logfile functionality was only added for three of the participants, as it was still in development when the first participants started.

At the end of the pilot, the participants were asked to fill in a questionnaire, to collect data about the user experience. This questionnaire consisted of 24 statements that the participants had to rate on a 7-points Likert scale. The statements are about four different topics: effectiveness, messages, usability, and lay-out. Furthermore, participants could contact the researchers with questions or problems during the pilot. All the feedback given by the participants at the start, during and at the end of the pilot, is considered as feedback.

There were seven participants (four women, three men), with an age between 65 and 77 years old. Six of the participants filled in the questionnaire at the end of the pilot (three women, three men). One participant could not fill in the questionnaire due to time limits, however she did provide some general feedback.

### 4.3.2 Results pilot

#### 4.3.2.1 Interaction

The logfiles from the tablets of the three participants that had a version with logging, were analysed. The first logfile showed that the participant used the application for 24 days and received 1888 notifications during this period. The participant responded on 15 (0.8%) of these notifications directly. Furthermore, the participant had 127 interactions on her own initiative. It needs to be noted that these logs sometimes overlap (for example entering and cancelling some intake, and adding an extra item multiple times which might be by mistake). In general, the log showed an usage pattern in which the participant updated her diet intake on a few moments a day (periodic recording).

The second logfile showed that the participant used the application for 33 days and received 1622 notifications during this period. The participant responded to 176 of these notifications directly (11%). The user had interactions with the system on her own initiative 51 times.

The third logfile showed that the participant used the application only for 8 days. According to the logfile the application was not always on, as it was not sending notifications for the whole day. This participant received 187 notifications, to 22 of these there was a direct response (12%). The number of user interactions that were logged that were on the participants own initiative is 5.

Figure 4.2 shows the active and responsive interaction rates of the three logfiles. Active interaction means that the user interacted with the system on its own initiative, whereas responsive interaction means that the user reacted on a notification or reminder send by the system.

#### 4.3.2.2 Perception

Next to the logfiles, the scores on the questionnaires of the six participants have been analysed. In order to compare and combine the scores for the four aspects (effectiveness,

**Table 4.1:** Overview of positively and negatively evaluated topics with their mean score

| | **Positively evaluated topics** | **Negatively evaluated topics** |
|---|---|---|
| **Effectiveness** | Motivation/help diet compliance (4.3) | Necessary functions available (2.5) |
| | Accuracy of the data (4.8) | Lives up to expectations (3.3) |
| **Messages** | Gentle tone (5.2) | Motivating (3.2) |
| | Not a compelling tone (5.2) | Interesting (2.8) |
| | Relevant for personal situation (4.5) | Problems the users experience (3.2) |
| | Believable/trustworthy (4.7) | Connection with perception and age (3.3) |
| **Usability** | Easy to use (4.2) | Irregularities (2.8) |
| | User friendly (4.2) | Would recommend the application (3.5) |
| | Easy to get along with (5.5) | Nice to use (3.3) |
| | Content with application and tablet (4.3) | Continuing to use (2) |
| | | Pleasant to use (3.3) |
| **Lay-out** | Appealing (4.2) | Professional (3.8) |
| | Clear (4.5) | |

messages, usability, and lay-out), the interpretation for all the scores needed to be made the same (namely: a higher score is better).

All the scores on the questions are combined per aspect, Figure 4.2 shows a boxplot for each aspect. The boxplot shows that for all the aspects the average score is around 4, which means neutral. However, the boxplot also shows that the scores vary a lot among the participants, especially for the usability.

Table 4.1 shows the average score for each statement and which topics (derived from the statements) have been evaluated positively (average $\geq 4$) and which topics have been evaluated negatively (average $< 4$). Overall, 12 topics are evaluated in a positive way and 12 topics are evaluated in a negative way. Next to the questionnaires the participants also gave extended feedback. All comments that have been made regarding the application that can improve the user experience in the next design, are taken into account as feedback. An important remark that has been made by participants is that the system is not flexible enough, the eating times are for example the same for each participant and there are only limited options to enter your diet when you do not follow the diet advises exactly. Statements from the questionnaire that have been evaluated negatively are that the messages are not motivating and interesting. In the feedback, one of the participants mentioned that new facts about protein or the research, send in the messages every day, would motivate to use the application.

### 4.3.3 Lessons learned

The pilot has been very useful to identify possible pitfalls, and different user profiles. Based on the logfiles, two different usage profiles are identified: notification driven usage and active recording. In Figure 4.2, Participant 5 shows the active recording profile, characterised by a high percentage of active interaction, Participant 6 and Participant 7 show a notification driven profile, in which the user has a high rate of responsive interaction. Since there are only three logfiles available to analyse, it might be that there are more usage profiles.

A comment that was made in the pilot is that the eating times were too fixed. Furthermore, it was noted that it differs per diet advice which eating moments are used. Therefore, the next version needs to have more flexibility in showing the diet plans. There was also a lack of flexibility in how participants could enter their (protein) consumption. During the pilot it became clear that the diet advises had more options for each eating moment. Before designing the next version of the application it was therefore important to clarify, in collaboration with the dietitians, how the diet plans for the next study will be set up.

In the refined design the application needs to have a greater added value. As mentioned by a participant, the application was just a reminder to eat and a digital version of the diet from the dietitian. The statements of the questionnaire that have to do with continuing to use the application and the availability of all the desired functions scored, on average, the lowest.

Furthermore, it became clear that a careful re-evaluation of all the messages and texts, so that they match the perception and age of the participants is important. It is valuable to look at ways to make the messages more suiting the experiences of the users to make them more interesting and motivating.

## 4.4 System redesign

Since the set up of the diet plan in the final study changes, and as a response to the lessons learned, the function requirements of the system are refined in Section 4.4.2. Section 4.4.1 will briefly elaborate on the new diet plan structure. The final design will be discussed in Section 4.4.3.

### 4.4.1 Diet plans for final study

The diet plans for the final study need to be followed for around six months, while in the pilot this was only three weeks. The diet plans will be designed in a more flexible way, which is also in accordance with the lessons learned during the pilot. The dietitians create a diet plan set up in which people need to have an intake of a certain number of points for each eating moment (comparable to the system of Weight Watchers[1]). A product can be worth 0 - 8 points (with steps of 0.5), indicating a category of the grams of protein in the product. The goal of this system is to make it more easy for participants to variate their diet. They will still be provided with an example menu based on their own preferences, but this is just a guideline. Besides this they will receive a protein table containing various products and product categories, including the number of protein points they are worth per portion of the product. Furthermore, the protein list will contain a sign indicating which products are so called energy products, which means they contain more calories than average products.

The dietitians stressed the importance of using both household measures (e.g. a cup or a spoon) as well as exact measures (weight or volume) for the input of the products a participant has eaten. The preference for one of these measures differ per participant. This contributes to the flexibility of the system.

[1] `https://www.weightwatchers.com/nl/`


**Figure 4.3:** Screenshots of the final design. Top left: the main screen with the overview of the eating moments and the consumed protein points; top right: a personalised notification; bottom left: window to fill in the protein points; bottom right: window to adapt the current menu.

Participants are also instructed to keep their intake at the requested level for each eating moment of the day. It is strictly advised to compensate for a meal where less than the requested level of proteins is consumed. However, it is not a big problem to consume too much protein now and then.

### 4.4.2 Extended function requirements

Some of the functional requirements described in Section 4.3 are refined as follows:

- Req. 1 is refined as: The system should give an overview of the personalised diet plan, per day, with the focus on the protein intake for each eating moment, expressed in protein points. There are sub requirements added:
  - The application must contain a suggested menu for each eating moment.
  - The suggested menu is first based on the diet plan of the dietitian, but during the study these suggestions will get tailored to the user.
- Req. 2 is extended with the warning to not send notifications too often (i.e. not every five minutes as in the prototype).
- The req. 4 and 5 are combined into one requirement: The user can enter the number of protein points for each eating moment. This can either be done directly or by entering the consumed products.

Furthermore, the following requirements are added to the list of requirements:

8. The diet plan should be loaded in a dynamic way (i.e. the names, times and number of moments), to make sure that it can be tailored to every user.
9. The system should take the diet rules into account e.g. the protein points.
10. The user should be able to enter his/her meal using either exact measures (grams and millilitres) or household measures (e.g. cups or teaspoons).
11. The user gets visual feedback on its intake compared to the requested intake.
12. The user should be able to also enter intake on another day than today.

With these extended functional, the aim is to improve the added value of the application. This is mainly done by adding the functionality to enter all the products of a meal in the system and giving the user feedback according to the diet rules.

### 4.4.3 Refined design

In the refined design each eating moment has a threshold value, which is the recommended number of proteins for that advice. Furthermore, a suggested menu is available for each eating moment, that adapts to the preferences of the user (req. 1). The bottom right of Figure 4.3 shows such a suggested menu (on the left). The adaption is done as follows: if the user changes something in the suggested menu, this is stored and when this change has been made three times, the change will become part of the suggested menu. Names, notification times and the number of eating moments are now adapted to each specific user (req. 8), the top left of Figure 4.3 shows a part of a diet plan for a day.

The notifications and reminders (req. 2) are still send in the same way as in the prototype, but the user now has only two options (see the top right of Figure 4.3): enter the intake (req. 4/5) or enable a reminder (req. 3). Notifications are kept on the screen for around 15 minutes and a user receives 3 notifications per hours at most (req. 2).

Moreover, the messages are further tailored based on scores on the Ten Item Personality Measure (TIPI) questionnaire [10].

In accordance with the literature from Section 4.2.1 the participant receives a *consensus messages* when he/she scores low on *conscientiousness* or *openness*, *commitment messages* when he/she scores high on *conscientiousness*, and *authority messages* in all other cases.

New messages are written, keeping in mind the abovementioned Cialdini principles. Furthermore the messages are written both in a formal and informal way, which way participants prefer is asked during their intake. Two different framings are used to personalise for the gender of the participant: gain framing for female participants and the loss framing for male participants. The politeness strategy *questions* is used to make the messages polite and persuasive.

The user has the ability to add its intake in the number of protein points (see the bottom left side of Figure 4.3) or by entering the complete meal. When entering the meal the user can use suitable household measures or exact measures to indicate the amounts (req.

10). For showing alternative products to the user the diet rules are followed (req. 9). The products from the protein list are categorised in order to provide useful orderings in the application.

In Section 4.2.2 it was noted that it is important to have a quick and easy administration of the intake. When designing the system this has been taken into account in multiple ways. The user has two ways to enter the intake; directly via the points (see the bottom left of Figure 4.3), and by entering all the products consumed (req. 4/5), this is also called the meal adviser. The suggested menus adapt to the user's preferences, and lists of products to choose from get a shortlist with the users top five products.

To meet req. 11, scales are added to each eating moment to indicate the user's consumption for that eating moment, as well as for the whole day (see Figure 4.3). Each scale contains an indicator with the requested intake and, if the user has some intake for that moment or day, the scale is coloured to indicate this intake. The scales for the eating moments are coloured green when the user had the right intake, orange when the intake was 0.5 points less/more then the target, and it is red in all other cases. At the bottom of the home screen (see the top left of Figure 4.3), users have the ability to change the day they are entering their intake for up to three days ago and three days ahead (req. 12).



## 4.5 First evaluation of refined design

After implementing the refined design, the new version of the application is evaluated with four participants from the pilot, during individual interviews. The main goal of the interviews is to see whether the feedback received during the pilot is successfully used to improve the prototype.

During the interview, the participants are asked to work with improved system. Together with the researcher the meal adviser is opened via a notification shown on the screen of the tablet. The participants are asked/explained how to select an alternative and how to add an item to the meal. The researcher assists the participants if needed and explains some of the features of the application (e.g. adaptation to the user and visualisation of the protein intake). The interaction between participants and the system is observed.

The participants are also asked to answer to some of the questions from the first questionnaire. Since the participants only have a short time to work with the application not all questions are relevant. Furthermore, open questions are asked about the improvements and the willingness to use the application.

### 4.5.1 Results from the interviews

Despite the low sample size, quantitative data from the questionnaires is analysed. Figure 4.4 shows the averages of the scores in the questionnaire of the prototype, both when including all questions in for this topic as well as only including the questions that have also been asked in the follow-up questionnaire. All topics are now scored with an average that is above the mean score of 4, which means all topics are evaluated positively. The highest average is for the lay-out. Open questions were also asked within the interviews. The comments made within these interviews are used to support the findings on all the

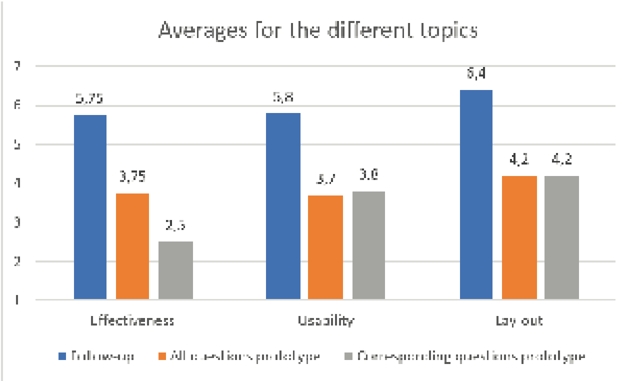



**Figure 4.4:** Graph showing the averages of the two questionnaires on different topics

topics from the questionnaires, as well as to get feedback on topics that were not in the questionnaire. In order to analyse those, they were combined based on their topic first and summarised in this analysis.

Figure 4.4 shows that all the topics have increased with at least two points. The highest increase (3.25) is for the effectiveness. This increase is also found for the question about whether participants wanted to use the application, which was slightly adapted from the question in the original questionnaire. The question about whether the participants could easily get along with the application had the smallest increase (0.5).

Participants further noted that the application had improved. The application was evaluated as more flexible, mainly due to the fact that you can enter your consumption more specifically and for multiple days at once. It was also noted that the lay-out was clean and professional. The adaptivity in the application was seen as a good addition that could also make it easier to work with the application. Finally, participants indicated that they would use the application because it provides an easier way to enter their daily intake. It was noted that it might motivate users with a lack of knowledge to help them with their diet. However, a participant noted that he might use the application less often once he was familiar with the diet. Another participant noted that she would not use the notifications, but would instead register on specific moment(s) of the day.

The importance of good instructions was stressed by multiple participants during the interviews. They suggested a more extended instruction compared to the instruction with the prototype, as well as a good instruction manual.

## 4.6 Discussion and future work

The most important lesson learned during the development of the system is that both the system as well as the participants need to have flexibility. The system needs to be flexible in how the diet plans of the participants are written. The participants need to have the freedom to enter what they have eaten. The system enables them to enter their intake with multiple measures (protein points, household measures and exact measures). The efforts made to improve this flexibility are evaluated positively during the follow-up evaluation. However, there might be some limitations. The protein table might be limiting for users, as not all possible products are covered by this table, but it is more on a general level. Furthermore, the categorisation that is made for the products might not be the most optimal categorisation for each user. Products can only be found in one category, however they might suit multiple categories. Furthermore, the household measures are now set to specific exact measures, while this might be personal (e.g. different cup sizes ). In a future version of the application more effort can be put in ways to make the search for a specific product more efficient and into personalisation of the measures used.

This pilot research was focused on the interaction that participants had with the system and the perception of the system of the participants. With the data collected in the current pilot it is not possible to draw any conclusions about the effect of the application on the compliance of the participants. In future research, as part of a larger study in the PROMISS project, this application will be used by a larger number of participants for a longer period of time. Within this research it will become clear whether the refined design improve the satisfaction of the users with the system. Furthermore, it will be possible to draw conclusions on the influence of the application on diet compliance. However, both the increase number of participants, as well as the increased duration of the study, will also give new insights in the user experience and interactions. It is expected that with this study it is possible to draw more general conclusions on design principles for persuasive communication systems to improve diet compliance for elderly.

Next to the refined design described in this chapter, also a second version will be tested within the final study. This version will use gamification to increase the motivation of participants to stick to their diet plan. Which gamification techniques can be used to do so, and how this can be combined with the current system, will be investigated in another project within this research.

Beyond the scope of the existing project setup, further extensions of the system are also possible. The current version of the application uses simple thresholds to adapt the menu to each individual user. However, with the use of machine learning techniques the suggested menu items could be adapted more specific to not only the user but also to for example the day of the week or other eating choices made during the day. Using machine learning could also improve how efficient the users can navigate through the lists of possible alternatives or extra items, as it could create a smart fast choice menu instead of simply showing the most frequently chosen products on top. It might be possible to use the data gathered during the large experiment that will be conducted in the near future for designing such algorithms.

## Bibliography

[1] Promiss. `http://www.promiss-vu.eu/`, accessed: 2018-09-27

[2] Al-Razgan, M.S., Al-Khalifa, H.S., Al-Shahrani, M.D., AlAjmi, H.H.: Touch-based mobile phone interface guidelines and design recommendations for elderly people: A survey of the literature. In: Int. Conf. Neural Inform. Processing. Springer (2012)

[3] Bentley, F., Tollmar, K.: The power of mobile notifications to increase wellbeing logging behavior. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1095–1098. ACM (2013)

[4] Carter, M.C., Burley, V.J., Nykjaer, C., Cade, J.E.: Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. Journal of medical Internet research **15**(4) (2013)

[5] Chittaro, L.: Tailoring web pages for persuasion on prevention topics: Message framing, color priming, and gender. In: Intl. Conf. on Persuasive Technology. pp. 3–14. Springer (2016)

[6] Cordeiro, F., Epstein, D.A., Thomaz, E., Bales, E., Jagannathan, A.K., Abowd, G.D., Fogarty, J.: Barriers and negative nudges: Exploring challenges in food journaling. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. pp. 1159–1162. ACM (2015)

[7] Coughlin, S.S., Whitehead, M., Sheats, J.Q., Mastromonico, J., Hardy, D., Smith, S.A.: Smartphone applications for promoting healthy diet and nutrition: a literature review. Jacobs journal of food and nutrition **2**(3), 021 (2015)

[8] Fogg, B.J.: Persuasive technology: using computers to change what we think and do. Ubiquity **2002**(December), 5 (2002)

[9] Goldberg, L.R.: The structure of phenotypic personality traits. American psychologist **48**(1), 26 (1993)

[10] Gosling, S.D., Rentfrow, P.J., Swann Jr, W.B.: A very brief measure of the big-five personality domains. Journal of Research in personality **37**(6), 504–528 (2003)

[11] Hammer, S., Lugrin, B., Bogomolov, S., Janowski, K., André, E.: Investigating politeness strategies and their persuasiveness for a robotic elderly assistant. In: International Conference on Persuasive Technology. pp. 315–326. Springer (2016)

[12] Hellman, R.: Universal design and mobile devices. In: International Conference on Universal Access in Human-Computer Interaction. pp. 147–156. Springer (2007)

[13] Hickson, M.: Malnutrition and ageing. Postgrad. Medical Journal **82**(963) (2006)

[14] IJsselsteijn, W., De Kort, Y., Midden, C., Eggen, B., Van Den Hoven, E.: Persuasive technology for human well-being: setting the scene. In: International conference on persuasive technology. pp. 1–5. Springer (2006)

[15] Isanejad, M., Mursu, J., Sirola, J., Kröger, H., Rikkonen, T., Tuppurainen, M., Erkkilä, A.T.: Dietary protein intake is associated with better physical function and muscle strength among elderly women. British Journal of Nutrition **115**(7), 1281–1291 (2016)

[16] Johnson, R., Kent, S.: Designing universal access: web-applications for the elderly and disabled. Cognition, Technology & Work **9**(4), 209–218 (2007)

[17] Kobayashi, S., Asakura, K., Suga, H., Sasaki, S.: High protein intake is associated with low prevalence of frailty among old japanese women: a multicenter cross-sectional study. Nutrition journal **12**(1), 164 (2013)

[18] Robert, C.: Influence: The psychology of persuasion. William Morrow and Company, Nowy Jork (1984)

[19] Thomas, R.J., Masthoff, J., Oren, N.: Adapting healthy eating messages to personality. In: Int. Conf. on Persuasive Technology. pp. 119–132. Springer (2017)

Chapter 5

# Integrating gamification into a system to improve diet compliance for elderly users

This chapter is published as: **van der Lubbe, L.M.**, & Klein, M.C.A. (2020). Integrating gamification into a system to improve diet compliance for elderly users. In Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good (pp. 150-155).



**This chapter contains:**



## Abstract

Nowadays, gamification is applied in many areas, including healthy lifestyle promotion. In earlier work, a system has been proposed to stimulate diet compliance and adherence of participants of a trial within the PROMISS-project. In this chapter, we describe the design of a gamified version of this system. The goal of the gamification is to further stimulate

diet compliance and adherence to the system, but also to increase the knowledge about the diet and make the use of the system more fun. To do so, we implemented gamification elements (profile page, achievements, mini-games, and a reward garden) to address multiple behaviour change techniques. Based on a small evaluation, the system has been improved so that it can be used by participants of the PROMISS trial. At the end of this chapter, future improvements are suggested in the future work section.

## 5.1 Introduction

In an earlier work, a system used within the PROMISS-project was described, which aims to prevent malnutrition among elderly (Chapter 4, [11]). This system is used within a larger study towards a protein rich diet for elderly, which is part of the PROMISS-project. The system mainly aims to improve the diet compliance and adherence of a subset of participants of this study. The system uses persuasive communication to achieve its goal. Gamification – adding game elements to an existing task, in order to make it more motivating [14] – could further enrich the system and contribute to this goal. This chapter describes how different game elements are used to enrich the existing system for the PROMISS trial.

Gamification is applied in many different areas, among which is supporting a healthy lifestyle. Since more people of different ages own smartphones or other mobile devices, with their own stores for applications (apps), the number of apps to promote an healthy lifestyle has increased [13, 9]. The increasing number of 'internet of things'-applications, and the increased interest in preventive healthcare will further increase the popularity of such apps.

Earlier work described the basic system (without game elements) for the PROMISS-trial (Chapter 4, [11]). The current chapter describes the design of the gamification for this system. In the next section, a literature overview of gamification will be given, as well as an overview of systems with a comparable goal or target group. In Section 5.3, the design of our gamification is explained. This chapter will conclude with a summarizing conclusion and a future work section.

## 5.2 Related literature

This section explains in more detail what gamification is and how it can be achieved. Furthermore, it gives an overview of gamification of similar apps and the use of gamification for elderly users.

### 5.2.1 Gamification and game elements

Gamification can be described as *"the intentional use of game elements for a gameful experience of non-game tasks and contexts"* [14]. Foursquare[1], founded in 2009, is a

[1] https://foursquare.com/

successful and early example of the use of gamification. Foursquare added game elements to their location-based social network to encourage people to use their service. Even before Foursquare, gamification in the form of loyalty programs was used by organisations to create consumer loyalty.

Recently, gamification has gained more and more attention, both in commercial fields as well as in academia, as it has many opportunities in the rising number of digital applications. Gamification is nowadays applied in a wide variety of contexts, e.g. education, health, commerce, and sustainable consumption [3].

Game elements play an important role in gamification. These game elements, or also called game mechanics, form the motivational factors that are at the core of gamification. Examples of such mechanics are points, leaderboards, badges, and achievements [22].

Game elements are also used in another type of application with a serious purpose, a serious game: *"a game in which education (in its various forms) is the primary goal, rather than entertainment"* [12].



### 5.2.2 Apps for a healthy lifestyle

There are different types of apps in field of healthy lifestyle. The majority of the apps found in a review paper of gamification for health and well-being [6] focused on physical activity and (healthy) nutrition. The review mentions seven promises of gamification for health and well-being, and argues to what extent these promises are covered by the reviewed studies. The first promise is that it can intrinsically motivate the users, as games can do. However, in their review it was found that studies often look at behavioural measures, and gamification is mainly used as a positive reinforcement. Intrinsic motivation is not studied. The second promise is that it has a broad accessibility through mobile technology and ubiquitous sensors. Although these techniques are used in the reviewed apps, it is not researched if there are any differences with stationary delivery modes. The third and forth promises are that gamification has a broad appeal and a broad applicability. There is a large and broad target audience for health apps using gamification. Moreover, gamification is and can be applied to many different aspects of health. The fifth and sixth promises are the cost-benefit efficiency and the fit to everyday life. However, the reviewed paper do not study this. The seventh promise is that gamification supports well-being. Positive effects for this are found in different researches.

The following game elements, in order of frequency (high to low), were found in the review of gamification for health and well-being [6]: rewards, avatars, leaderboard, social interaction, levels, progress, story/theme, challenges, and feedback. It was found that apps often include different game elements. It is therefore not possible to determine the effect of single game elements. Overall, the conclusion of this review is that gamification for health and well-being apps has a positive effect on the affect, behaviour, cognition, and user experience. Often, it also has a positive effect on the user's health and/or well-being. When the latter effect is not found, this is often due to the context or the way in which gamification is used, or due to a mismatch of the game elements and the target group.

Apps for healthy lifestyle often have the goal to change their users' behaviour. In order to do so, behaviour change techniques (BCTs) can be used. A list of 26 BCTs is constructed by Abraham and Michie (2008) [1]. Reviews towards health interventions via websites showed that interventions including more BCTs are more effective [19, 20]. In a review of apps to improve healthy lifestyle of children and adolescents, it was found that an app uses, on average, 6 out of the 26 BCTs. According to this is comparable to the number of BCTs (4 - 8) found in reviews towards healthy lifestyle apps for adults. The most popular BCTs (found in $\geq$ 40% of the reviewed apps) that were found in apps for children and adolescents were: instructions, general encouragement, contingent rewards, feedback on performance, self-monitoring behaviour, and social comparison [13]. In many of the games within this review gamification is used.

It is found that the number of BCTs in an app is positively associated with the quality of an app [13], as well as on engagement and information scores but not for functionality and aesthetics scores. To measure this quality of health apps, the Mobile App Rating Scale (MARS) can be used [18]. Using this, it is found that the number of app features is only positively associated with the total and the engagement MARS score. Moreover, a positive correlation was found between the number of features and the number of BCTs. The reviewed apps score the highest on functionality, aesthetics, and engagement. However, they score lower on information. This could mean that the focus of developers is mostly on creating easy to use and functional apps [13].

The review of [9] is focused on health and fitness. Again, it was found that many games include gamification elements to motivate the users. Most used gamification element is social or peer pressure, followed by digital rewards, competitions, leaderboards, and levels of achievement or rank. In some apps it was found that real world prizes were used as game element. Next to gamification elements, elements that are more typical for games are also found. These include, among others, feedback, reinforcement, self representation with avatars, time pressure and narrative contexts. Moreover, behaviour measures such as self-monitoring, goals setting and peer pressure are found in the applications.

Overall, it has been found in multiple reviews that health apps often use gamification to achieve a behavioural goal with the user. A broad audience is reached by such games, as well as a large domain of different aspects of the health domain. The apps often use game elements such as rewards, leaderboards, social aspects and challenges. These game elements are used to include BCTs in the apps, which contributes to the quality of the app. The social aspects of apps do not only contribute to the integration of BCTs in apps, it is also a way to advertise apps to new users [16].

### 5.2.3 Gamification for elderly

Gamification can be and is used for elderly users, however this target group needs some special attention in the design process. For example, it is found that game elements can work differently for younger people compared to how they work for elderly [8]. Also, the ease of use of gamification declines with age [7]. Usability is important to keep in mind when designing for elderly. Using guidelines about motivational factors, such as making it fun and clearly defining the benefits of it, is important when designing for elderly [2]. In the design of the app itself, it is also important to keep in mind guidelines for elderly users. In the previous paper about this system (Chapter 4, [11]), an overview of such guidelines is given. Remarks have been made about the need for larger screens, with bigger fonts and spacings to ensure readability for an older target group. The navigation of the system should be clear and reduce or avoid drag-and-drop interactions. Moreover, it is important to avoid an overload of information on one page, as well as scrolling.

Although the target group of elderly users needs some special attention, gamification is used in applications designed for them, but also serious games for elderly exist. Serious games for elderly are promising in the field of prevention and rehabilitation [21]. For example, many different games exist for fall prevention, such as [17, 15]. However, serious games and gamification is also used in different other areas relevant to elderly users. For example, there exists a serious game for elderly to train their verbal resilience to doorstep scams [10], and a gamified training for the cognitive functioning of elderly, in the context of safe driving [4]. As these examples show, gamification can be targetted at elderly users, if the design is tailored to their needs.

## 5.3 Design of gamified version



This section explains the gamified system. First, the structure of the design process is explained. Next, the different steps from the design process, as well as the final gamified design, are explained.

### 5.3.1 Design process

The starting point of the gamified app is an existing system designed for the PROMISS-project. The design process starts with designing the goal(s) and determine which game elements suit the goal(s). Next, a first design of the gamification was implemented. To evaluate this version, an interview has been held with one person from the aimed target group. Moreover, feedback from dietitians and experts from the PROMISS-project is gathered. All feedback is used to make some changes after which the gamified version is used within the trail. More about this evaluation is explained in 5.3.4.

In the rest of this section, the goal of the gamification is discussed, then the final design of the gamified version, and finally the lessons that have been learned from a prototype evaluation.

### 5.3.2 Goal of gamified version

As stated in Section 5.2.1 gamification is the addition of game elements to an existing task. Therefore, it is important to first think about the goal of the gamification, the desired game elements that can achieve this, and how this can suit the existing task. In this case, the task to gamify is a food diary with the guidelines from the PROMISS-project diet included.

The goal of the gamification is trifold. On the one hand, the gamification should further enhance the compliance with the dietary advice and the adherence to using the app. As discussed in Section 5.2.2, gamification is often used in health related apps to increase the motivation of the user. By adding game elements to the app, the hypothesis is that the users will enjoy using the app more, and be more motivated to use the app. Moreover, game elements can be used to reward users for their diet compliance.


**Figure 5.1:** Example of profile page

Secondly, it is of interest whether the gamification also improves the knowledge of the users. Users of the initial system passively work with the protein point system, the system calculates the points each meal is worth. With gamification it is possible to make users more actively work with the protein point system, which could enhance their knowledge. Finally, games are meant to be fun. The third goal is therefore to make the app (more) fun to use.

When thinking about gamifying the protein point system, one problem arises: users should not feel motivated to consume as many protein points as possible. This is represented in the colour scales used for the protein points and thresholds for each eating moment and the day total. However, it also needs to be taken into account in the mechanics of the gamified version.

The final gamified system consists of four components: a profile with usage statistics, achievements, mini-games and a reward garden. In Section 5.3.3 the functioning and rationale behind each component is explained. The lessons learned from the evaluation of the prototype are described in Section 5.3.4.

### 5.3.3 Gamified system

At the end of the section, Table 5.1 summarises how each game element is connected to the goals explained in Section 5.3.2 with a rating from 0-5 stars. Zero stars meaning no expected contribution, five stars meaning a very strong expected contribution. Moreover, Table 5.2, shows how different BCTs, based on the review of [13], are used within the app, again with the five star system.

#### 5.3.3.1 Personal profile

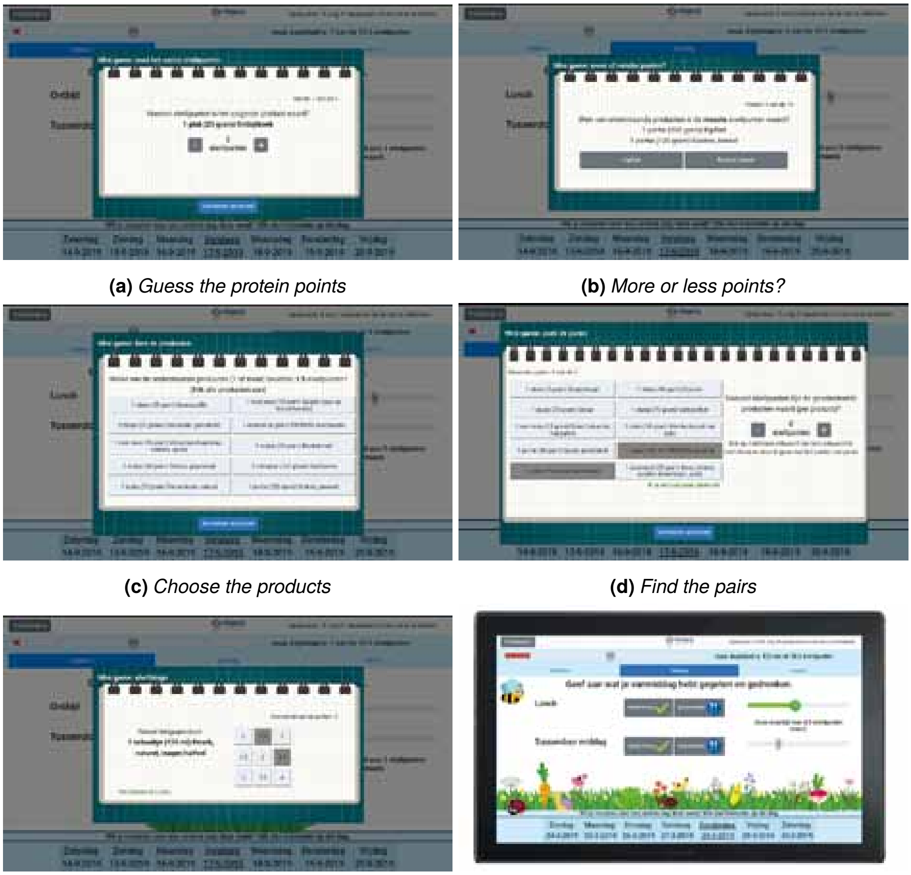
Figure 5.1 shows a profile page for a user. On top of the profile page the user is greeted. This is personalised to the users preference for a form of address; informal or formal.

This is followed by a list of statistics of their usage of the app:

- Start date and number of days since the start;
- Total number of protein points registered and accuracy[2];
- Total number of days for which some proteins are registered and accuracy, see Footnote 2;
- Total number of eating moments for which protein points are registered and accuracy, see Footnote 2.

[2]The accuracy shows the percentage of the expected registration that is registered by the user.





**(a)** *Guess the protein points*

**(b)** *More or less points?*

**(c)** *Choose the products*

**(d)** *Find the pairs*

**(e)** *Protein bingo*

**(f)** *Final stage reward garden*

**Figure 5.2:** Screenshots of the five mini-games and reward garden

| | | Goals | | | |
|---|---|---|---|---|---|
| | | Diet compliance | App adherence | Knowledge | Fun |
| Game elements | Profile page | *** | ***** | | * |
| | Achievements | ** | ***** | | *** |
| | Mini-games | **** | **** | ***** | ***** |
| | Reward garden | | ***** | | **** |

**Table 5.1:** Overview of game element and the addressed goals

The profile furthermore lists the achievements that the user has achieved, together with the date on which the player received it. Moreover, the next achievements in each category is shown. The achievements are explained in the next section.

The profile page contributes to the self-monitoring of the behaviour of the user, see Table 5.2. The profile summarises the user's performance over the full duration of the app, while they can only see their detailed performance for a few days. Moreover, the accuracy confronts users with their adherence directly. It is expected that the profile will mostly contribute to the adherence to the app, as it mainly shows the user statistics about this. However, it will also contribute to the diet compliance, as the statistics also give insight in this. Finally, to some users the profile can be seen as fun, as they like to follow their usage and loyalty to the app and diet.

#### 5.3.3.2 Achievements

Consecutive achievements (42 in total) are created for three different categories: protein points (18 achievements), days of registration (15), and registration of eating moments (9). They have been designed in such a way that users can receive them during the full duration of the study, with a wide variety of description texts. When a player earns a new achievement, a notification is shown accompanied by a cheering sound effect.

Achievements give general encouragement, provide feedback on the performance and provide contingent rewards to the player, see Table 5.2. Players can be motivated by the fact that they can collect achievements. As is shown in Table 5.1, this mainly contributes to the app adherence, because the number of days of registration and the number of moments registered are awarded. It does not say much about the adherence to the diet, as the achievements are only related to the total number of protein points. Moreover, achievements will be perceived as fun to accomplish.

#### 5.3.3.3 Mini-games

Five different short quiz-like mini-games, have been created for the gamified version. They are based on a subset of the products in the protein table used within the PROMISS trail; namely all specific products. General products (such as 'meat, fish, vegetarian component') are excluded from this subset. All games randomly choose products from this set of products, however some games have constraints on these choices, which are explained below. With each game the user can earn up to maximum 10 so-called 'game points'.

| BCT | App element |
|---|---|
| Provide information about behaviour-health link | Messages * (only some strategy) |
| Provide information on consequences | Messages * (only some strategy) |
| Provide information about others' approval | Messages * (only some strategy) |
| Provide general encouragement | Achievements ***<br>Coloured intake ** |
| Provide instruction | Diet advice composer ** |
| Model/demonstrate the behaviour | Diet advice composer * |
| Prompt specific goal setting | Thresholds diet **** |
| Prompt self-monitoring of behaviour | Food diary functionality *****<br>Profile page *** |
| Provide feedback on performance | Coloured intake *****<br>Achievements *****<br>Mini-games **** |
| Provide contingent rewards | Achievements *****<br>Mini-games ****<br>Reward garden **** |
| Stress management | Reward garden * |

**Table 5.2:** Overview of relation of BCTs and app elements

These points are used to improve the reward garden, which is described in the next section. In each game sound effects are used to indicate right or wrong answers.

To make the games suitable for elderly users, the games have an easy navigation (e.g. no drag and drop actions or moving objects) and are text-based rather than relying on complicated visual components. Furthermore, there is no time pressure, so that players can read the texts and play the game at their own pace. Moreover, the games are easy to understand and play, which aims to reduce the threshold for people to play the games.

The mini-games serve as a reward for meeting the threshold for an eating moment. As discussed earlier, consuming more than the recommended number of proteins is not necessarily better and therefor not stimulated. After filling in its intake for an eating moment on the right level, the user is invited to play one of the five games (chosen at random). It is also possible to skip the game. An explanation is given at the beginning of each game. The printed manual that users received also includes an explanation for each game, as well as an explanation of the other gamification elements.

The mini-games are the only game element that serves all different goals, as shown in Table 5.1. As it is a reward for registering a meal with the advised number of proteins, it stimulates both diet compliance as well as app adherence. This is also a way to give feedback on the performance of the user, as indicated in Table 5.2. Moreover, it stimulates the players' knowledge about the protein points, as all games are questioning this knowledge. This resulted in easy and quiz-like games, giving the users a way to both increase and show their knowledge on the protein points. Finally, the games are meant as a fun element within the application. The following paragraphs briefly discuss the different mini-games.

**Guess the protein points** In this game there are five rounds in which the player is shown one product and its standard portion size. The player is asked how many protein points this portion is worth (see Figure 5.2a). Products can only be chosen by the system once per game. The player can indicate the number of protein points with a plus and minus button as (s)he is used to from entering its diet. The player earns two points for a correct answer and one point when the answer is 0.5 protein points off.

**More or less points?** This game consists of 10 rounds of one question: 'Which product contains more/less protein points?'. For each game, it is randomly chosen whether the question will be about *more* or *less* points. The first product of each question is randomly chosen. The second products needs to be worth a different number of points, within a 1.5 point range from the first product. The player is shown two products and their standard portion and is asked to choose the right answer (see Figure 5.2b). For each correct answer the player earns one point.

**Choose the products** In this game the player is shown 10 products and their standard portion. The player needs to click all products with a specific number of points (see Figure 5.2c). The list consists of 10 unique products, chosen randomly. The number of points that the player needs to find is also chosen randomly, but needs to appear at least once in the list of 10 products. Each product that is in the right category (clicked or not clicked) is worth one point.

**Find the pairs** In this game the player is shown 10 products and their standard portions. The player is asked to make five pairs of products with the same number of protein points. For this game a different subset of products is used, namely the set of products that can form a pair, so only point categories with at least two products are included. A pair should consist of two different products and for each number of points only one pair can appear in the list.

For each correctly formed pair the player receives one point. The player can try as many times as (s)he needs. The player therefor always earns a minimum of five points. When a correct pair is formed, the player is asked how many points the products are worth (separately). If this question is answered correctly, the player receives one extra point for each pair. In Figure 5.2d an example is shown of a formed pair and the bonus question.

**Protein bingo** In protein bingo the player starts with 10 points and an empty bingo card. On this bingo card the following numbers are shown: 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, and 4 (see Figure 5.2e). The player is shown a product and its standard portion. This product is chosen randomly, but products are only used once during the game and only products worth a number of points that is still available on the bingo card are selected. The player then needs to choose the right number of protein points by selecting an unused number on the bingo card. When the player chooses an incorrect number, one point is reduced from the player's points total for this game. When the player has no more numbers left on its bingo card, (s)he wins the remaining number of points. When there are no more points left, the game is over.

#### 5.3.3.4 Reward garden

By playing the mini-games, the player earns game points which also improves its "reward garden". At the start, the background of the application is blank. The more game points are earned, the more of the background is filled garden elements. On the top right of the screen, the total number of game points is shown, as well as the next threshold for receiving an improvement. The garden has 37 stages. The first improvement is given after earning five game points, which is possible by playing one game. Each subsequent stage requires between 50 and 150 game points. If the player improves its garden (s)he will receive a notification, accompanied by the cheering sound that is also used for the achievements. Figure 5.2f shows the final stage of the reward garden.



The reward garden is used to increase the app adherence, as it mainly rewards playing the mini-games, and is meant as a fun aspect in the app (see Table 5.1). Previous research argued that using a garden in a game, either to look at or to do activities in, can have a stress reducing effect [5]. Next to reducing stress, the reward garden mainly provides a contingent reward (see Table 5.2).

### 5.3.4 Lessons from evaluation the prototype

As the core functionality of the system was not changed by the gamification elements, and to limit the time of the design phase, no additional pilot study has been performed. Instead, one participant (female) from the first pilot study (Chapter 4, [11]) was invited to play the different mini-games and see the profile page with an explanation of the different gamification elements. Overall, she was positive about the gamification elements and she was enthusiastic about the rewarding garden. However, from her interaction with the mini-games it turned out that she quickly attempted playing the games, without fully reading the instructions. She also reported this. In the prototype the instructions were provided when the game started. To increase the attention to the instructions, they are moved to a separate screen at the start of each game. Participants have to click a button to move from the instruction page to the game. Moreover, the participant had some remarks about elements that were unclear to her. To address this, some explanations or visual elements in the games are changed or added.

The new version of the app is also shown to dietitians and experts from the PROMISS-project that were not involved in the development process. They were very enthusiastic about the mini-games as well and liked how it tests the knowledge about the protein point system. They mentioned that they would have liked such games during their education to learn about nutritional values.

## 5.4 Conclusions and future work

Gamification is getting more attention in many fields, among which is the domain of healthy lifestyle promotion. This chapter described how an existing system is extended with gamification elements to fulfil different goals: increase diet compliance, app adherence, knowledge about the diet, and as a way of making the usage of the app more fun. As earlier work suggests, using multiple BCTs in an app increases its potential. Therefore, different BCTs are addressed by the different game elements used within this gamified system.

By taking into account both the goals of the system, the requirements from the study, and design guidelines for elderly users, a design has been made for a gamified system, consisting of the following four components: profile with usage data, achievements, mini-games, and a reward garden. From an evaluation of the first design of this system with one of the earlier participants from a pilot study, it became clear that the visuals and instructions of the system were not clear enough to the user. Adjustments have been made so that the instructions could not be overlooked and visual indications of changes in the game were added.

The system is currently used within the large clinical trial. Both app usage data will be collected, as well as data about the usability of the system and the gamification elements. The overall appreciation of following the diet of all participants will be compared to see if the gamification has an effect on this.

The design of this system showed that on the one hand, it is important to match the goals with the chosen gamification elements when designing for a specific purpose. On the other hand, the target group requires to match the chosen gamification elements with their special needs. This has let to a simple, yet diverse gamified system, with multiple game elements using different BCTs.

### 5.4.1 Future work

The results of the currently running larger study are expected halfway 2020. It is expected that this results will give more insights in the usability of such a system in a diet program for elderly. This could result in new design guidelines and/or best practices, which can be used in future researches towards similar types of systems.

Moreover, in future research it can be interesting to test other variations on the current design. The mini-games in this design are mainly text-based. It is yet unknown whether these games would have a different effect and evaluation when they would include more visual components. Moreover, the mini-games are relying on random choices of products. However, it can be possible to create different difficulty levels for the games. This could be done manually, by determining products or product combinations that are harder to distinguish. It could however also be done with machine learning techniques by learning from the behaviour of users. Questions that are more often answers incorrectly could be classified as harder compared to questions that are often answered correctly.

Furthermore, it could be studied if achievements can be made adaptive to individual users. For example by ensuring that each next achievement is within a desirable distance in time of the current achievements so that players stay motivated.

# Bibliography

[1] Abraham, C., Michie, S.: A taxonomy of behavior change techniques used in interventions. Health psychology **27**(3), 379 (2008)

[2] De Carvalho, R.N.S., Ishitani, L., Nogueira Sales De Carvalho, R., et al.: Motivational factors for mobile serious games for elderly users. Proceedings of XI SB Games (2012)

[3] Hamari, J., Koivisto, J., Sarsa, H., et al.: Does gamification work?-a literature review of empirical studies on gamification. In: HICSS. vol. 14, pp. 3025–3034 (2014)

[4] Hiraoka, T., Wang, T.W., Kawakami, H.: Cognitive function training system using game-based design for elderly drivers. IFAC-PapersOnLine **49**(19), 579–584 (2016)

[5] Höchsmann, C., Walz, S.P., Schäfer, J., Holopainen, J., Hanssen, H., Schmidt-Trucksäss, A.: Mobile exergaming for health—effects of a serious game application for smartphones on physical activity and exercise adherence in type 2 diabetes mellitus—study protocol for a randomized controlled trial. Trials **18**(1), 103 (2017)

[6] Johnson, D., Deterding, S., Kuhn, K.A., Staneva, A., Stoyanov, S., Hides, L.: Gamification for health and wellbeing: A systematic review of the literature. Internet interventions **6**, 89–106 (2016)

[7] Koivisto, J., Hamari, J.: Demographic differences in perceived benefits from gamification. Computers in Human Behavior **35**, 179–188 (2014)

[8] Link, M.W., Lai, J., Bristol, K.: Not so fun? the challenges of applying gamification to smartphone measurement. In: International Conference of Design, User Experience, and Usability. pp. 319–327. Springer (2014)

[9] Lister, C., West, J.H., Cannon, B., Sax, T., Brodegard, D.: Just a fad? gamification in health and fitness apps. JMIR serious games **2**(2) (2014)

[10] van der Lubbe, L.M., Gerritsen, C., Formolo, D., Otte, M., Bosse, T.: A serious game for training verbal resilience to doorstep scams. In: International Conference on Games and Learning Alliance. pp. 110–120. Springer (2018)

[11] van der Lubbe, L., Klein, M.C.: Designing a system with persuasive communication to improve diet compliance for elderly users. In: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. pp. 234–241. ACM (2019)

[12] Michael, D.R., Chen, S.L.: Serious games: Games that educate, train, and inform. Muska & Lipman/Premier-Trade (2005)

[13] Schoeppe, S., Alley, S., Rebar, A.L., Hayman, M., Bray, N.A., Van Lippevelde, W., Gnam, J.P., Bachert, P., Direito, A., Vandelanotte, C.: Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. International Journal of Behavioral Nutrition and Physical Activity **14**(1), 83 (2017)

[14] Seaborn, K., Fels, D.I.: Gamification in theory and action: A survey. International Journal of human-computer studies **74**, 14–31 (2015)

[15] Silva, J., Oliveira, E., Moreira, D., Nunes, F., Caic, M., Madureira, J., Pereira, E.: Design and evaluation of a fall prevention multiplayer game for senior care centres. In: International Conference on Entertainment Computing. pp. 103–114. Springer (2018)

[16] Souza-Júnior, M., Queiroz, L., Correia-Neto, J., Vilar, G.: Evaluating the use of gamification in m-health lifestyle-related applications. In: New Advances in Information Systems and Technologies, pp. 63–72. Springer (2016)

[17] Stanmore, E.K., Mavroeidi, A., de Jong, L.D., Skelton, D.A., Sutton, C.J., Benedetto, V., Munford, L.A., Meekes, W., Bell, V., Todd, C.: The effectiveness and cost-effectiveness of strength and balance exergames to reduce falls risk for people aged 55 years and older in uk assisted living facilities: a multi-centre, cluster randomised controlled trial. BMC medicine **17**(1), 49 (2019)

[18] Stoyanov, S.R., Hides, L., Kavanagh, D.J., Zelenko, O., Tjondronegoro, D., Mani, M.: Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth and uHealth **3**(1), e27 (2015)

[19] Van Genugten, L., Dusseldorp, E., Webb, T.L., Van Empelen, P.: Which combinations of techniques and modes of delivery in internet-based interventions effectively change health behavior? a meta-analysis. Journal of medical Internet research **18**(6), e155 (2016)

[20] Webb, T., Joseph, J., Yardley, L., Michie, S.: Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. Journal of medical Internet research **12**(1) (2010)

[21] Wiemeyer, J., Kliem, A.: Serious games in prevention and rehabilitation—a new panacea for elderly people? European Review of Aging and Physical Activity **9**(1), 41 (2012)

[22] Zichermann, G., Cunningham, C.: Gamification by design: Implementing game mechanics in web and mobile apps. " O'Reilly Media, Inc." (2011)

# Chapter 6

# Insights in the effect and experience of a diet tracking application for older adults in a diet trial

This chapter is published as: **van der Lubbe, L.M.**, Klein, M.C.A., Visser, M., Wijnhoven, H.A.H., Reinders, I. Insights on the Effect and Experience of a Diet-Tracking Application for Older Adults in a Diet Trial. Technologies 2022, 10, 31.

6

**This chapter contains:**

## Abstract

With an ageing population, healthy ageing becomes more important. Healthy nutrition is part of this process and can be supported in many ways. The PROMISS trial studies the effect of increasing protein intake in older adults on their physical functioning. Within this trial, a sub-study was performed, researching the added effect of using a diet-tracking app enhanced with persuasive and (optional) gamification techniques. The goal was to see how older adult participants received such technology within their diet program. There were 48 participants included in this sub-study, of which 36 completed the study period of 6 months. Our results on adherence and user evaluation show that a dedicated app used within the PROMISS trial is a feasible way to engage older adults in diet tracking. On average, participants used the app 83% of the days, during a period of on average 133 days. User-friendliness was evaluated with an average score of 4.86 (out of 7), and experienced effectiveness was evaluated with an average score of 4.57 (out of 7). However, no effect of the technology on protein intake was found. The added gamification elements did not have a different effect compared with the version without those elements. However, some participants did like the added gamification elements, and it can thus be nice to add them as additional features for participants that like them. This chapter also studies whether personal characteristics correlate with any of the other results. Although some significant results were found, this does not give a clear view on which types of participants like or benefit from this technology.

## 6.1 Introduction

As populations are ageing, healthy ageing becomes more important. One of the aspects contributing to healthy ageing is healthy nutrition. Research has shown that smartphone applications (apps) used during dietary interventions have a higher adherence compared with websites or paper diaries [2]. Moreover, it is shown that smartphone apps can improve the diet compliance of participants, e.g., see [4]. Creating a specific app that meets the needs of the target group is preferred over using a general journal [3].

Gamification is used in many different domains, among which is healthy lifestyle [5]. Gamification can be described as “the intentional use of game elements for a gameful experience of non-game tasks and contexts” [13]. One of the main goals of adding gamification to an app is that it stimulates the motivation of users so that the outcomes of the application are increased. Although the ease of use of gamification declines with age [7], it can be used for older adults in different contexts; e.g., different rehabilitation or prevention games for older adults exist, for example, to make physical training to prevent falls more fun, e.g., see [14].

The PROMISS trial aims to increase protein intake in older adults with a relatively low protein intake [12, 11]. Good physical functioning is important for healthy ageing and living independently at home. In the 6-month PROMISS trial, the effect of protein on physical function is studied. Part of this trial was the persuasive technology sub-study, in which participants were provided with a tablet app and a so-called foodbox, both designed specifically for diet tracking within this trial (Chapter 4 and 5, [8, 9]). Persuasive communication and gamification techniques were used in the app to enhance adherence to food intake registration and diet compliance. The goal was to see how such technology was received by the older adult users and, as part of this assessment, whether it has an effect on, for example, the outcome of the dietary advice and experience with the diet.

The current chapter describes the outcomes of the persuasive technology sub-study of the PROMISS trial. First, in Section 6.2, a short overview of the background of this work is given. The research question, materials, and methods used are explained in Section 6.3. In Section 6.4, an overview of the participants is given, followed by analyses of different research questions. At the end of each subsection about a research question, a summary of the results can be found. Finally, the discussion (Section 6.5) summarises the results and discusses possible limitations and future possibilities. Finally, the most important lessons learned are highlighted in the conclusions in Section 6.6.

## 6.2 Background

In previous work, more details on the background and related work of the work presented in the current chapter were presented (Chapter 4 and 5, [8, 9]). This section gives a brief overview of the relevant background literature.



In the field of healthy lifestyle applications, different types of apps can be found, with the majority focusing on physical activity and (healthy) nutrition [5]. In the review by Johnson (2016), seven promises of gamification for health and well-being are mentioned. Among those promises are the broad appeal and applicability and the motivation it gives to users. Apps for a healthy lifestyle often aim to change the behaviour of the users. Achieving this is often more successful when behaviour change techniques (BCTs) are incorporated in the intervention [15, 16].

Previous research has shown that tailoring the design of an application to an older adult target group is good for, among other things, the effectiveness, efficiency, and user satisfaction [6]. Examples of guidelines that can be used when designing for older adults are using larger screens, fonts, and spaces between buttons [1]. In addition, the methods of interaction, navigation, and information loading are aspects that need tailoring to the target audience of older adults. In general, for all target groups, it is found that, in the domain of food intake registration, an easy and quick administration of food intake is important [4].

## 6.3 Materials and methods

The PROMISS trial addresses whether dietary advice to increase protein intake to $\geq$1.2 g/kg adjusted body weight (aBW)/day (this measure is explained in Table 6.1) is beneficial for physical functioning in community-dwelling older adults. Products are valued according to a protein point system designed for this trial. More details on the design of the trial can be found in its design paper [12]. The PROMISS app is designed using the same information and instructions on products and protein as all participants of the PROMISS trial receive in an information booklet.

The PROMISS app is provided to our participants on a dedicated tablet. The app is a diet journal, in which participants can register their intake in two ways: directly entering the protein value of the meal or choosing products to create a meal. In the latter case, the app calculates the protein value of the meal. The app can also help to replace products in a meal, keeping the protein value the same. Registering intake can be performed for each instance of eating in the participants' personal dietary advice. The app gives an overview of the protein value of each meal and the whole day and compares it with the personal thresholds in the dietary advice. Moreover, the application uses persuasive communication in its notifications that remind participants to eat. Next to the app, the foodbox is a custom-made box in which specific protein products provided from the trial, such as puddings and protein bars, can be stored. Taking them out of the box triggers a notification in the app that helps users register the protein value associated with the product. More details on the design of the system can be found in a previously published paper (Chapter 4, [8]). In a second variant of the app, gamification elements (rewards, achievements, a profile page, and mini-games) were added. These elements were added to study their added value for adherence, for increasing protein knowledge (via the mini-games), and for making the app more fun to use. More details on the design of the gamified variant can be found in a previously published paper (Chapter 5, [9]).

### 6.3.1 Participants

The persuasive technology sub-study was only performed in the PROMISS trial site in the Netherlands. This trial consisted of three study groups: a control group and two intervention groups, in which participants received personalised dietary advice. Recruitment and selection was part of the PROMISS trial and is described in [12, 11]. The study included community-dwelling older adults ($\geq$65 years) with a lower protein intake. Exclusion criteria, such as health concerns, are listed in Reinders et al. [12]. Participants in the intervention groups could opt-into the sub-study, as they need to obtain dietary advice that can be put in the app. Only one participant per household could be included in the sub-study because it was likely that participants living together influence their outcomes by working together. The sub-study shared the measure moments with the trial, namely baseline, 3 months, and 6 months. Participants were free to stop using the technology provided in the sub-study earlier, and the time at which instances were measured remained the same.

An overview of the participants in the trial in the Netherlands and the persuasive technology sub-study (PT) can be found in Figure 6.1. In total, 48 participants took part in our sub-study. Participants who did not fill in the persuasive technology evaluation questionnaire or who had $\leq$20 days of input on the tablet were excluded from the analysis and considered drop-outs. In total, 12 participants dropped out, of which 3 also dropped out of the trial. For this data analysis, the focus was on the 36 PT participants and the 41 participants who were not in the PT group (no PT). Five participants from the PT group had data input on the tablet for more than 200 days; those five participants were considered the COVID-19 extension group. The PT group contained 15 females and 21 males, with a mean age of 74 (SD = 4.7, minimum = 68, maximum = 85). The majority of the participants had an education level of secondary education or higher. There was no information available about the experience of participants with technology. There were 17 participants in the normal condition and 19 participants in the gamification condition. In total 26 participants received a foodbox. Participants could stop using the foodbox whenever they wanted, without becoming a drop-out.

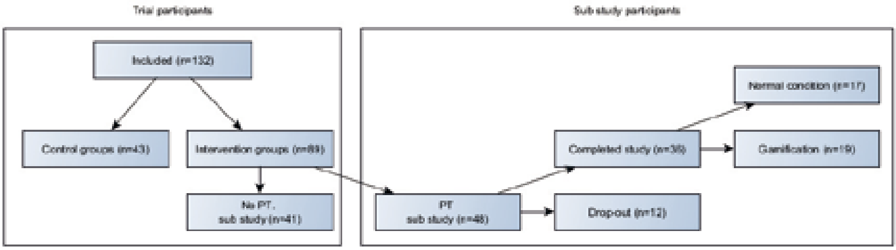



**Figure 6.1:** Overview of participants in the trial and sub-study. [10]

Statistical tests were performed to determine whether participants in the PT group have comparable characteristics with the no-PT participants at baseline. No significant results are found for age, sex, education level, and mean protein intake at baseline in the grams per kilogram-adjusted body weight per day. This means that the PT and no-PT groups were not significantly different. It is compared whether there were significant differences between participants in the normal and gamification conditions. Again, no significant differences were found; thus, it can be concluded that the participants in both conditions were comparable.



### Drop-out

As mentioned, twelve participants dropped out from the sub-study. Three of these dropped out of the main trial. For the other participants, remarks they made about dropping out were noted.

One participant returned the tablet as he/she did not see the added value of the tablet over using a paper diary. This participant also told us to not use (and that they are not familiar with) technology such as a tablet or smartphone. The lack of familiarity with technology and understanding of the tablet was a reason for dropping out for at least two other participants as well.

Two other participants noted that using the tablet was educational, but after some time, they no longer needed to use it or that it was time-consuming and bothersome. Moreover, finding products in the tablet was a problem for one of these participants, as the ordering was sometimes unclear and products were missing. Another participant also mentioned the tablet being time-consuming as a reason to drop out.

Another participant used the tablet once but told us to forget about it and therefore completely quit using the tablet. While another participant used the tablet on some days during the trial, but they did not use the tablet for the full trial for an unspecified reason. It needs to be noted that this participant also missed some appointments for the main study as well.

### 6.3.2 Design

The sub-study had two tablet conditions: the normal and the gamification conditions. The assignment to these conditions was semi-random because the inclusion for the gamification condition started later due to logistical reasons, and the aim was to keep the size of the groups equal. Aside from this, the foodbox was a storage box for protein products provided to some participants in the trial and can automatically register the consumption of these products. As not all participants received the products that can be registered by the foodbox and as the number of foodboxes available was limited, not all participants received one, regardless of their study condition.

The main research question of the sub-study was as follows:

> How is a gamified persuasive diet-tracking system as part of a diet program received by older adults?

To answer this question, different research tasks (RT) were defined:

1. To explore the **adherence** of participants who use the diet-tracking system.
2. To statistically test if using persuasive technology increases **protein intake**.
3. to statistically test if the **experience of the dietary advice** improves using persuasive technology.
4. To explore the **experience of the persuasive technology**.
5. To explore the effect of **gamification** on participants' **knowledge** about the amount of protein in food products.

### 6.3.3 Materials

Different data sources were used to address these research tasks. Table 6.1 shows the different data sources and for which RTs those sources are relevant. Within this study, the focus was to evaluate the current version of the application and its effects. Thus, mainly quantitative data in the form of questionnaire answers, protein intake, and log data were collected.

The measures for the dietary advice evaluation can be found in Appendix 6.A.1; these were designed within the larger PROMISS trial [12]. The persuasive technology questionnaire evaluates the experienced effectiveness and user-friendliness of the tablet and the foodbox, the notifications shown in the app and the gamification elements, using seven-point Likert scales. All questions of this questionnaire can be found in Appendix 6.A.2. This questionnaire was custom made for this study. The goal was to obtain more details on specific aspects of the system, and therefore, it was decided to custom make our questionnaires. Moreover, to ensure that older adults understand the wordings, the statements were made concrete about specific aspects. The structured questionnaires with (mainly) closed questions aimed to give insight into the evaluation of the current system. As the PROMISS trial was only performed once, more in-depth information for improvements was not gathered. In cases where statistical tests were used to complete a research task, a threshold of 0.05 was used for significance. In the tables in Section 6.4, significant results are marked with an asterisk (*).

**Table 6.1:** Overview of different data sources.

| Data Type | Explanation | Measured | Participant Group | RT |
|---|---|---|---|---|
| Tablet data | Logging of participant interaction with app (e.g., which buttons clicked, which products chosen). | During tablet use | Sub-study | 1, 2, 4, 5 |
| Mean protein intake in g/kg adjusted BW/d | Mean protein intake in gram per kilogram body weight per day, adjusted for the BMI of the participant. For participants with a low/high BMI, their protein intake was calculated based on a healthy BMI. | At baseline, 3 months, and 6 months | All | 2 |
| Dietary advice evaluation | General evaluation questions about how participants experienced/value the dietary advice. | At 6 months | All | 3, 5 |
| Persuasive technology evaluation | Questionnaire about user experience and expected effectiveness of the persuasive technology. | At 6 months | Sub-study | 4 |

The tablet collected different data, mainly to ensure correct functioning of the tablet. Every click was logged into the database (action and timestamp), with these data, we can analyse what participant clicked in the application. For each meal that was registered, the protein value and timestamp were saved, and if provided, the products and amounts were also added to this logging.

Due to the COVID-19 pandemic in 2020 [1], it was not possible to conduct final measurements for participants for some time. Due to this, 14% of the participants were able to use the tablet for an extended period. As a cut-off point, 200 days (normal maximum duration) of usage was assumed. Above that, this is marked as a 'COVID-19 extension'. Whether the fully extended trial or the trial period (of maximum 200 days) was used is indicated in

[1] https://time.com/5791661/who-coronavirus-pandemic-declaration/, accessed on: 31 November 2021

the analyses.

### 6.3.4 Procedure

Participants received their tablets and verbal instructions, together with written information, at the beginning of their diet trial period. After that, participants could use the tablet for six months, during which they could contact the researcher if they experienced technical difficulties. Table 6.1 shows which data were collected at which measure moment. All questionnaires were part of the procedure of the main trial.

## 6.4 Results

For each research task, data analysis and statistical testing (if applicable) were performed. Throughout this sections, summary boxes highlight the most important results that are found.

### 6.4.1 Adherence (RT1)

Figure 6.2 shows the duration of the trial for the PT participants, specifying the number of days in which they did and did not provide input. On average, participants used their tablets for 133 days (SD = 51.03, minimum = 27, maximum = 198). To measure the length of participants' tablet usage period, the first and last dates on which they registered any input were used. Figure 6.3 shows the duration of the trials of five participants with an extended trial period of >200 days. They used their tablet on average for 196.4 days (SD = 57.19, minimum = 101, maximum = 239). Moreover, Figure 6.3 shows that three of these participants were active users for most of the trial, while the others used the tablet less frequently but continued to use the tablet during the extended part of the trial.

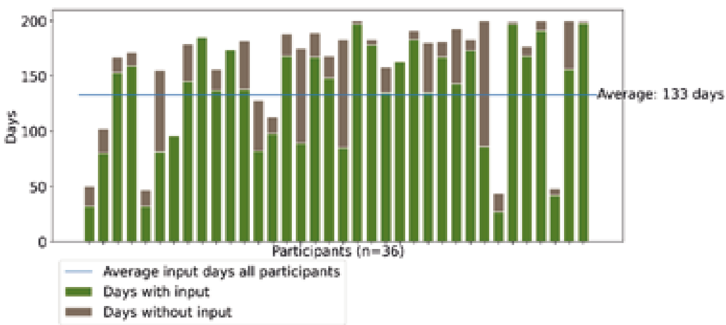



**Figure 6.2:** Number of days with and with tablet input in the trial per participant. [10]

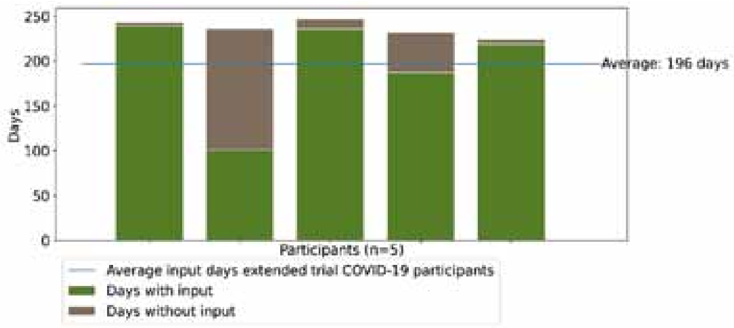



**Figure 6.3:** Number of days with and with tablet input in the trial per participant for the COVID-19 extension case. [10]

Figure 6.4 shows the division of days with and without inputs; for this figure, the full duration of the participants in the extended trial period was used, as this is a relative graph. On average, participants used the tablet 82.7% (SD = 16.60%, minimum = 43.80%, maximum = 100%) of the total tablet use days. Based on the average number of input days and the relative adherence, the adherence of participants was considered high.

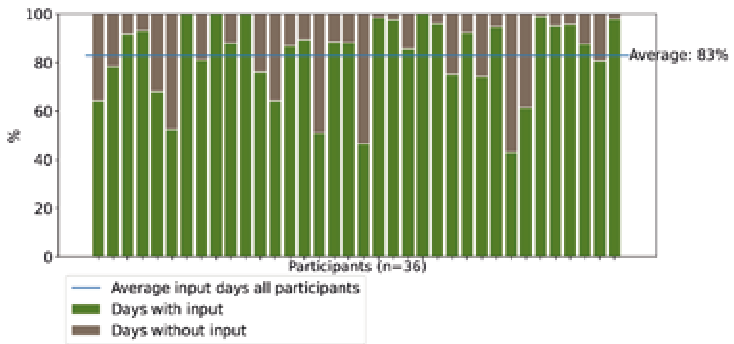



**Figure 6.4:** Percentage of (in)active days in the trial per participant. [10]

The adherence of participants is considered high:

- The average number of active days of participants without COVID-19 extension is 133 (SD = 51.03, minimum = 27, maximum = 198), see Figure 6.2.
- The average number of active days of participants with COVID-19 extension is 196.4 (SD = 57.19, minimum = 101, maximum = 239), see Figure 6.3.
- The relative number of input days for participants is 82.7% (SD = 16.60%, minimum = 43.80%, maximum = 100%), see Figure 6.4.

#### 6.4.1.1 Differences between tablet conditions for adherence

When studying the differences between the normal condition and the gamification condition (excluding COVID-19 extension) for the adherence, a slight difference in the mean duration length is found; see Figures 6.5 and 6.6. However, a *t*-test shows that this difference is not significant (*p*-value = 0.378).

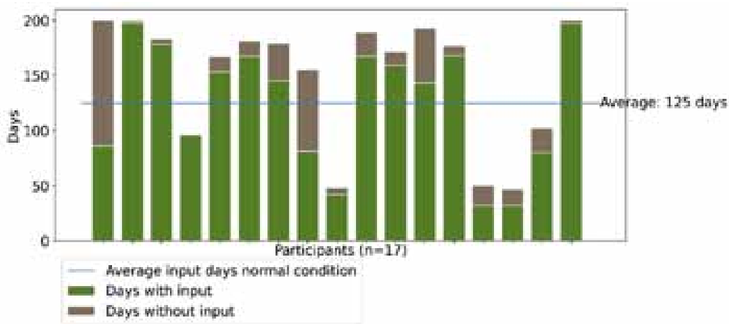



**Figure 6.5:** Number of days with and with tablet inputs in the trial per participant—normal condition.

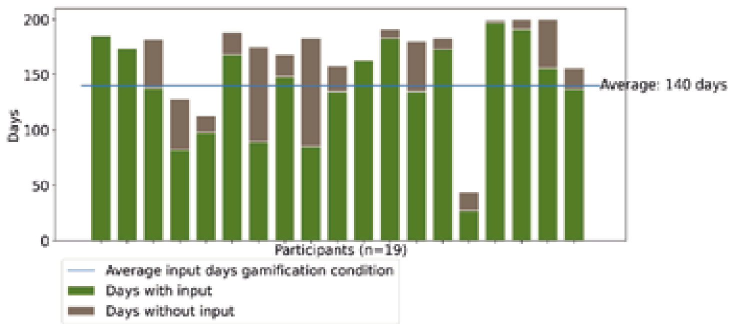



**Figure 6.6:** Number of days with and with tablet inputs in the trial per participant—gamification condition.

The relative graphs (including the COVID-19 extension), Figures 6.7 and 6.8, for the two conditions show that the average relative active usage is almost the same, namely 82% (SD = 17.03%) for the normal condition and 83% (SD = 16.67%) for the gamification condition. This difference in significance is tested with a *t*-test. The difference is not significant (*p*-value = 0.927). Based on these results, it is concluded that the version of the app did not influence the adherence of participants.

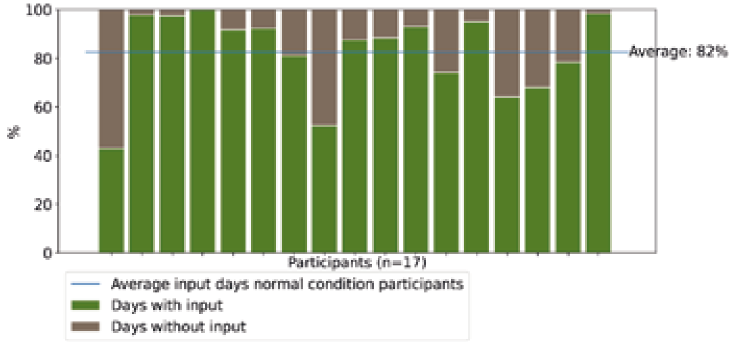



**Figure 6.7:** Percentage of (in)active days in the trial per participant—Normal condition.

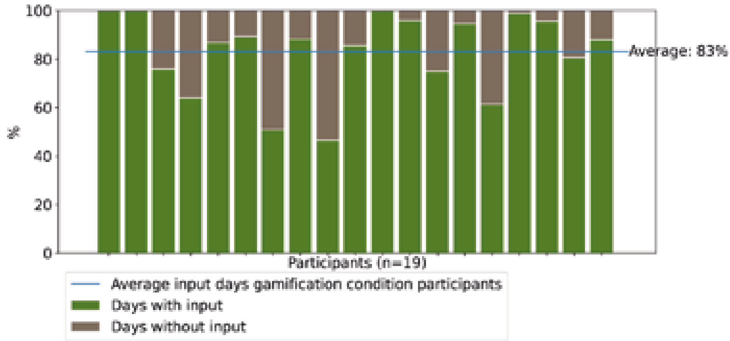



**Figure 6.8:** Percentage of (in)active days in the trial per participant—Gamification condition.

The version of the tablet application does not influence the adherence of participants:

- The average number of active days of participants (excluding COVID-19 extension) in the normal condition is 124.9 (SD = 56.94, minimum = 32, maximum = 198), see Figure 6.5.
- The average number of active days of participants (excluding COVID-19 extension) in the gamification condition is 140.2 (SD = 45.45, minimum = 27, maximum = 197), see Figure 6.6.
- The *t*-test for the difference between active days for the normal and gamification conditions is not significant (*p*-value = 0.378).
- The relative number of input days for normal-condition participants (including COVID-19 extension) is 82.4% (SD = 17.03%, minimum = 42.8%, maximum = 100%), see Figure 6.7.

- The relative number of input days for gamification-condition participants (including COVID-19 extension) is 83% (SD = 16.67%, minimum = 46.4%, maximum = 100%), see Figure 6.8.
- The *t*-test for the difference between the relative number of input days for the normal and gamification conditions is not significant (*p*-value = 0.927).

#### 6.4.1.2 Differences by personal characteristics for adherence

It is also statistically tested whether differences in adherence are caused by personal characteristics. It is tested whether the adherence of male and female participants is significantly different with a *t*-test. Both the mean number of days and the relative adherence are not significantly different (*p*-value = 0.990 and *p*-value = 0.725). With the help of the Pearson correlation, it is determined whether age or protein intake at baseline is correlated with the adherence of participants, both for the days of input and relative adherence. Both age (*p*-value = 0.254 and *p*-value = 0.904) and mean protein intake at baseline (*p*-value = 0.599 and *p*-value = 0.433) do not show significant differences. Finally, a one-way ANOVA is performed to test whether there are significant differences in the days of input or relative adherence between the education levels. Again, there are no significant differences found (*p*-value = 0.529 and *p*-value = 0.099). Based on all these statistical tests, it is concluded that personal characteristics do not influence the adherence of participants.

The personal characteristics of participants do not influence the adherence of participants. Different statistical tests showed no significant differences caused by personal characteristics (sex, age, protein intake at baseline, and education levels).

### 6.4.2 Persuasive technology and protein intake (RT2)

Table 6.2 shows the *p*-values from *t*-tests comparing the protein intake of the PT group with the protein intake of the no-PT group. Moreover, it shows whether the mean of all participants included in our sub-study (PT + drop-out) is different from the no-PT group, based on *t*-tests.

When comparing the PT participants with the no-PT participants, no significant differences in their protein intake were found. However, when comparing all the participants who were included in the sub-study with the no-PT group, significant differences for the intake at 6 months and the change between baseline and 6 months were found: the no-PT group has a higher change. It is however unclear what could explain this difference. If there would be an effect of the intention to treat, it is expected that this effect is also found for the active participants.

To investigate this further, the correlation between the number of input days in the tablet and the diet outcomes is studied. Table 6.3 shows the results of these analyses. For these analyses, the full duration of the trial, including the extension due to COVID-19, is taken into account. Again, only a significant correlation for the intake at 6 months and the change between baseline and 6 months for all included sub-study participants is found.

Figure 6.9 shows the percentage of participants in the PT and no-PT groups that reached the goal of the main trial 1.2 protein intake in grams per kilogram-adjusted body weight per day for the 3-month measurement. The difference between how many participants reached the goal in each group is not significant ($p$-value = 0.490).

Figure 6.10 shows this for the 6-month measurement. Again, the differences between the two groups are not significant ($p$-value = 0.621). In both groups, the number of participants who stuck with the goal protein intake decreased at 6 months compared with at 3 months. However, the change in the group of tablet users was smaller compared with the other participants of the intervention groups.

**Table 6.2:** Results $t$-tests for protein intake (expressed in g/kg adjusted BW/d) between the PT and no-PT groups. * $p < 0.05$.

| Measure | Mean PT | Mean All Sub-Study Participants | Mean No PT | *p*-Value PT | *p*-Value All Sub-Study |
|---|---|---|---|---|---|
| Mean protein intake—3 months | 1.2 SD = 0.24 (n = 34) | 1.2 SD = 0.27 (n = 43) | 1.3 SD = 0.23 (n = 38) | 0.194 | 0.175 |
| Mean protein intake—6 months | 1.2 SD = 0.19 (n = 36) | 1.2 SD = 0.23 (n = 44) | 1.3 SD = 0.28 (n = 38) | 0.086 | 0.019 * |
| Change protein—between baseline and 6 months | 0.4 SD = 0.25 (n = 36) | 0.4 SD = 0.27 (n = 44) | 0.5 SD = 0.30 (n = 38) | 0.132 | 0.026 * |
| Change protein—between 3 and 6 months | 0.01 SD = 0.28 (n = 34) | −0.02 SD = 0.29 (n = 42) | 0.05 SD = 0.26 (n = 37) | 0.549 | 0.230 |
| Change protein—between baseline and 3 months | 0.4 SD = 0.24 (n = 34) | 0.4 SD = 0.24 (n = 43) | 0.5 SD = 0.25 (n = 38) | 0.351 | 0.236 |

**Table 6.3:** Results from the Pearson correlation test between days with tablet registration and protein intake (expressed in grams per kilogram-adjusted body weight per day). * $p < 0.05$.

| Measures | All Participants Sub-Study | | PT Participants | |
|---|---|---|---|---|
| | Pearson Correlation Coefficient | *p*-Value | Pearson Correlation Coefficient | *p*-Value |
| Mean protein intake—3 months | 0.17 (n = 43) | 0.289 | 0.21 (n = 34) | 0.231 |
| Mean protein intake—6 months | 0.42 (n = 44) | 0.005 * | 0.21 (n = 36) | 0.217 |
| Change protein—between baseline and 6 months | 0.32 (n = 44) | 0.033 * | 0.11 (n = 36) | 0.529 |
| Change protein—between 3 months and 6 months | 0.16 (n = 42) | 0.316 | −0.09 (n = 34) | 0.626 |
| Change protein—between baseline and 3 months | 0.11 (n = 43) | 0.482 | 0.06 (n = 34) | 0.722 |

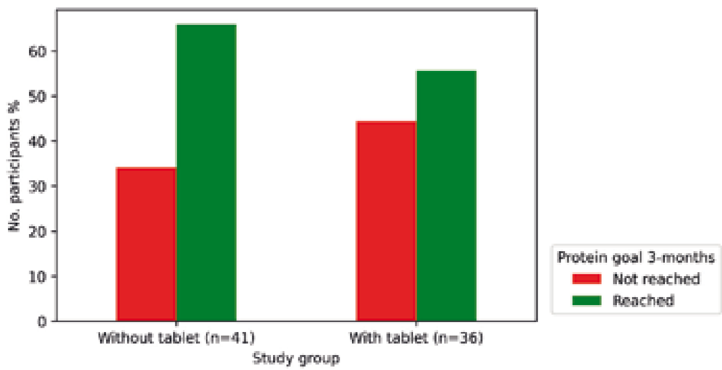



**Figure 6.9:** Participants reaching 1.2 protein intake in grams per kilogram-adjusted body weight per day at 3 months. [10]

We also analysed whether there was a difference between the groups when looking at participants who reached their protein goal at both instances of measured; see Figure 6.11. In the group without a tablet, half of the participants reached the goal at both moments and half did not. In the group with the tablet, this was more divided and the majority did not reach the goal at both moments. However, the differences between the two groups are not significant (*p*-value = 0.521).

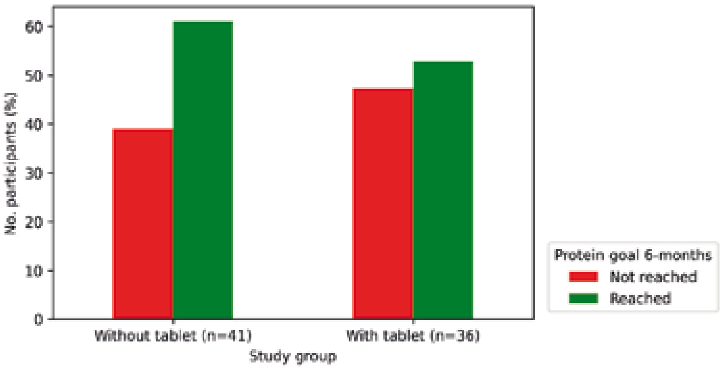



**Figure 6.10:** Participants reaching 1.2 protein intake in grams per kilogram-adjusted body weight per day at 6 months. [10]

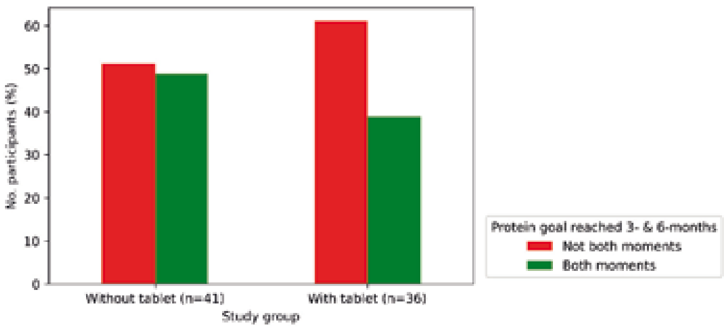



**Figure 6.11:** Participants reaching 1.2 protein intake in grams per kilogram-adjusted body weight per day at 3 and 6 months.

Using the persuasive technology did not increase the protein intake of participants:

- T-tests were performed to see if PT participants (n = 36) reached better diet outcomes compared with the no-PT participants (n = 41) and to compare all included sub-study participants (n = 48) with the no-PT participants; see Table 6.2. Significant results are found for all sub-study participants and the no-PT participants for the mean intake at 6 months (*p*-value = 0.019) and changed between baseline and 6 months (*p*-value = 0.026).
- Pearson correlations were calculated for the protein intake and the days with tablet registration for all sub-study participants and PT participants; see Table 6.3. Sig-

nificant correlations were found for mean intake at 6 months (*p*-value = 0.005) and change intake between baseline and 6 months (*p*-value = 0.033) for all sub-study participants.

- The result of the Chi²-test for the relation between the groups (PT and no PT) and reaching a threshold value of 1.2 g/kg adjusted BW/d at 3 months was not significant (*p*-value = 0.490).

- The result of the Chi²-test for the relation between the groups (PT and no PT) and reaching a threshold value of 1.2 g/kg adjusted BW/d at 6 months was not significant (*p*-value = 0.621).

**Table 6.4:** Resulting *t*-tests for protein intake—per condition (expressed in grams per kilogram-adjusted body weight per day).

| Measure | Mean Normal Condition | Mean Gamification Condition | *p*-Value |
|---|---|---|---|
| Mean protein intake in g/kg adjusted BW/d—3 months | 1.2 SD = 0.20 (n = 15) | 1.3 SD = 0.27 (n = 19) | 0.244 |
| Mean protein intake in g/kg adjusted BW/d—6 months | 1.2 SD = 0.23 (n = 17) | 1.3 SD = 0.15 (n = 19) | 0.302 |
| Change protein in g/kg adjusted BW/d—between baseline and 6 months | 0.4 SD = 0.29 (n = 17) | 0.4 SD = 0.22 (n = 19) | 0.418 |
| Change protein in g/kg adjusted BW/d—between 3 months and 6 months | 0.04 SD = 0.31 (n = 15) | −0.01 SD = 0.27 (n = 19) | 0.633 |
| Change protein in g/kg adjusted BW/d—between baseline and 3 months | 0.4 SD = 0.21 (n = 15) | 0.4 SD = 0.26 (n = 19) | 0.331 |

#### 6.4.2.1 Differences between tablet conditions for protein intake

For the PT participants in the normal and gamification conditions, the diet outcomes were compared using *t*-tests. The results of these tests are summarised in Table 6.4. No significant results are found, so the tablet condition did not influence the change in protein intake of participants.

The version of the tablet application does not influence the protein intake of participants:

- T-tests to see if participants from normal conditions had different changes in protein intake than participants from the gamification condition (Table 6.4) show no significant results.

#### 6.4.2.2 Differences by personal characteristics for protein intake

It is interesting to see whether the personal characteristics of participants in the PT group influenced protein intake. For the different measures of protein intake, statistical tests were performed: a *t*-test for sex, Pearson correlation coefficients for age and protein intake at baseline, and a one-way ANOVA for education level. The *p*-values resulting from these tests can be found in Table 6.5.

**Table 6.5:** *p*-values for protein intake (expressed in grams per kilogram-adjusted body weight per day) and personal characteristics. * $p < 0.05$.

| | ***p*-Value** | | | |
|---|---|---|---|---|
| **Measure** | **Sex** | **Age** | **Protein Baseline** | **Education** |
| Mean protein intake in g/kg adjusted BW/d—3 months | 0.886 | 0.100 | 0.052 | 0.219 |
| Mean protein intake in g/kg adjusted BW/d—6 months | 0.959 | 0.235 | 0.393 | 0.766 |
| Change protein in g/kg adjusted BW/d—between baseline and 6 months | 0.561 | 0.105 | $1 \times 10^{-5}$ * | 0.113 |
| Change protein in g/kg adjusted BW/d—between 3 months and 6 months | 0.957 | 0.036 * | 0.036 * | 0.058 |
| Change protein in g/kg adjusted BW/d—between baseline and 3 months | 0.661 | 0.269 | 0.192 | 0.648 |

The protein intake at baseline correlates significantly with the change between baseline and 6 months. To see if this is due to the tablet, the same test for the no-PT participants was conducted. This also showed a significant difference (*p*-value = 0.045), so we cannot conclude that this is an effect of the tablet. Moreover, age and protein intake at baseline correlate significantly with protein change between 3 and 6 months. However, when studying the data in more detail, it was found that one protein change is much higher compared with others. When removing this outliers, the statistical significance disappears. It, therefore, cannot be concluded that age or protein intake at baseline influences the effect of tablet use on protein intake.

The personal characteristics and tablet use of participants did not influence the protein intake of participants:

- Statistical tests for personal characteristics and protein intake (Table 6.5) show no significant differences for personal characteristics and tablet use.

### 6.4.3 Experience of the diet (RT3)

After 6 months, the dietary evaluation questionnaire (see Appendix 6.A.1) was used to assess the experience of users with the dietary advice. The PT group was compared with the no-PT group to see if there were significant differences for these questions. The results are shown in Table 6.6. No find significant differences were found for any of the measures. However, for both the usefulness of the dietary advice and the ease of finding products, the sub-study participants had a slightly better appreciation. The ease of finding products was one of the goals of the app.

**Table 6.6:** Results of the statistical tests for diet evaluation (range 1–5).

| **Measures** | **Means PT** | **Means No PT** | ***p*-Values** |
|---|---|---|---|
| Rating dietary advice (range 1–10) | 8.4 SD = 1.03 (n = 36) | 8.5 SD = 1.04 (n = 37) | 0.608 |
| Usefulness dietary advice | 4.2 SD = 0.65 (n = 36) | 4.1 SD = 0.70 (n = 37) | 0.475 |
| Extend to which dietary advice is followed | 4.3 SD = 0.61 (n = 36) | 4.4 SD = 0.49 (n = 37) | 0.087 |
| Ease to stick to dietary advice | 3.9 SD = 0.86 (n = 36) | 4.0 SD = 0.70 (n = 37) | 0.143 |
| Intend to continue to follow dietary advice | 3.9 SD = 0.67 (n = 36) | 4.0 SD = 0.71 (n = 37) | 0.093 |
| Ease to find products with similar protein amount | 3.9 SD = 0.85 (n = 36) | 3.8 SD = 0.71 (n = 37) | 0.402 |

Using persuasive technology does not seem to affect the experience of the diet:

- Statistical tests comparing the PT group with the no-PT group (Table 6.6) show no significant differences in the experience of the diet.

#### 6.4.3.1 Differences between tablet conditions for the experience of the diet

The same analyses were performed to see if there were differences in the experience of the diet between the two tablet conditions. The results can be found in Table 6.7. A significant difference is found for the rating of the diet, which is higher for participants in the normal condition.

**Table 6.7:** Results of the statistical tests for diet evaluation—per condition (range 1–5). * $p < 0.05$.

| Measure | Mean Normal Condition | Mean Gamification Condition | *p*-Value |
|---|---|---|---|
| Rating dietary advice (range 1–10) | 8.8 SD = 0.90 (n = 17) | 8.0 SD = 1.04 (n = 19) | 0.036 * |
| Usefulness dietary advice | 4.3 SD = 0.69 (n = 17) | 4.2 SD = 0.63 (n = 19) | 0.336 |
| Extend to which dietary advice is followed | 4.2 SD = 0.56 (n = 17) | 4.3 SD = 0.67 (n = 19) | 0.314 |
| Ease to stick to dietary advice | 4.0 SD = 0.71 (n = 17) | 3.9 SD = 0.99 (n = 19) | 0.500 |
| Intend to continue to follow dietary advice | 3.8 SD = 0.75 (n = 17) | 4.1 SD = 0.57 (n = 19) | 0.079 |
| Ease to find products with similar protein amount | 3.8 SD = 0.97 (n = 17) | 4.0 SD = 0.74 (n = 19) | 0.266 |

Different statistical tests compared the normal and gamification condition (Table 6.7). A significant difference was found for rating their diets (*p*-value = 0.036); the rating was higher for the normal conditions. Other aspects were not significantly different.

#### 6.4.3.2 Differences by personal characteristics for the experience of the diet

It was tested whether the personal characteristics of participants in the PT group influenced their evaluation of the application. For the rating, the same statistical tests as those used for the personality characteristics tests described in Section 6.4.2.2 were used. For the other measures, the Mann–Whitney U test was used for the sex characteristic and the Spearman correlation was used for the other characteristics. One significant difference was found: female tablet users score the question about continuing to follow the dietary advice significantly higher compared with male participants ($p$-value = 0.029). For the no-PT users, this difference was not found ($p$-value = 0.417). It, therefore, seems that the tablet has a higher effect on female participants' intentions to continue to follow dietary advice.

Different statistical tests for the associations between personal characteristics and the evaluation of the diet were performed. A significant result for intention to continue to follow dietary advice and sex ($p$-value = 0.029) was found, and the score was higher for women. Other differences associated with personal characteristics were not found.

### 6.4.4 Experience persuasive technology (RT4)

This section analyses how participants experience persuasive technology. To do so, the results from the evaluation questionnaire as well as some data on the interaction that participants make with the system were analysed. For the evaluation, differences between the two tablet conditions and the effects caused by personal characteristics were analysed.

#### 6.4.4.1 Evaluation questionnaire persuasive technology

Figure 6.12a shows the scores of the *experienced effectiveness of the tablet*. This shows that most participants find the tablet helpful (Q1) and the functionalities sufficient (Q2). However, the participants are more divided and less positive about how the tablet met their expectations (Q3). Participants are on average slightly negative about continuing to use the tablet after the study finished (Q4). One of the participants added to this that he/she was now familiar with the protein diet. Another participant added to their answer to Q4 that it would be nice to use the tablet from time to time, instead of every day. Overall, concerning the experience effectiveness (Q1–Q4), participants were slightly positive (mean = 4.6, SD = 1.94).

Figure 6.12b shows the scores for the *user-experience of the tablet*. The only question that scored below neutral was the question about irregularities (Q6). During the trial, different irregularities were found, some of which could not be solved during the trial. Overall, the user-friendliness was evaluated slightly positive, with an average score of 4.9 (SD = 1.99) for Q5–Q10.

Figure 6.12c shows the scores for the *foodbox questions* (Q11–Q13), which were on average evaluated negatively (mean = 2.8, SD = 1.79). Only the questions about ease to use (Q12) were evaluated a bit more positive.

There was some confusion among participants about what was meant by *notifications*. Moreover, for Q18, one participant answered both 4 and 5; in this case, a score of 4.5 was used as an exception. One participant added to the questions that he/she did not use the reminders but filled in his/her consumption whenever it suited him/her. Most statements were evaluated neutral and with diverse scores; see Figure 6.12d. For the friendly tone (Q14), the score was a bit higher. In general, the notifications (Q14–Q21) were evaluated with an average score of 4.4 (SD = 1.72).

The messages that participants received could be of one of three styles, depending on their personality. It turned out that only two styles were used: COM (30 participants) and AUT (6 participants). The COM-style has no personality requirements, which means that most participants did not have a personality that suited a specific communication style.

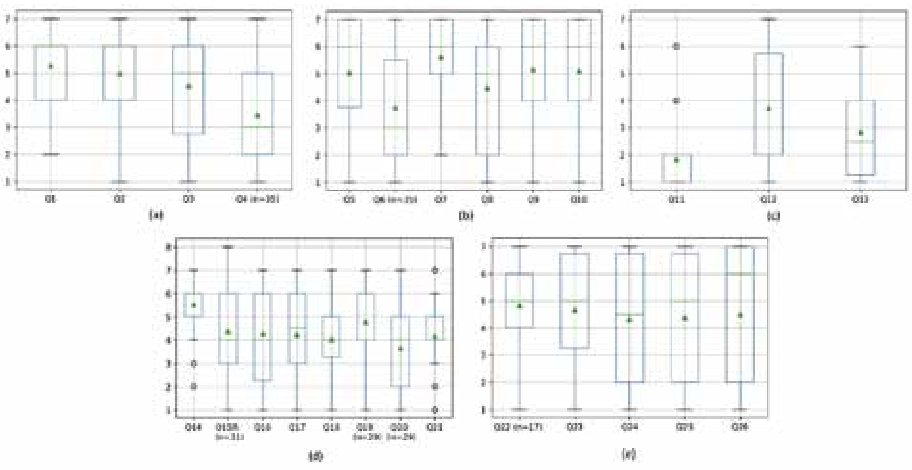


**Figure 6.12:** Boxplots of the PT evaluation questionnaire per topic for all active participants—scale 1-7 [10], Triangles: mean scores; Circles: outliers; 1 = strongly disagree (negative evaluation), 7 = strongly agree (positive evaluation); the questions can be found in Appendix 6.A.2. (**a**) Experienced effectiveness tablet (n = 36). (**b**) User-friendliness of the tablet (n = 36). (**c**) Foodbox (n = 26). (**d**) Notifications (n = 30). (**e**) Gamification (n = 18).

Table 6.8 shows the results from the statistical tests for the means of the two styles. Participants with the COM style were slightly more positive about the messages they received, mainly about the friendliness of the messages. Although for individual questions no significant differences were found, a significant difference was found when combining all of the scores. The COM style was appreciated significantly better compared with the AUT style.

The evaluation of the gamification elements was only relevant for participants in the gamification tablet condition. Figure 6.12e shows that the scores about *gamification* (Q22–Q26) were very divided and, on average, slightly above neutral (mean = 4.5, SD = 2.25). For the mini-games, participants were the most positive about the learning aspect of the games.

**Table 6.8:** Average scores of the notifications for different communication styles—scale 1–7, 1 = strongly disagree, 7 = strongly agree. * $p < 0.05$.

| | **Both Styles** | **AUT Style** | **COM Style** | ***p*-Value** |
|---|---|---|---|---|
| Friendly tone | 5.5 SD = 1.31 (n = 30) | 5.0 SD = 2.00 (n = 5) | 5.6 SD = 1.15 (n = 25) | 0.389 |
| Compelling (reversed) | 4.3 SD = 1.84 (n = 31) | 4.6 SD = 1.14 (n = 5) | 4.3 SD = 1.9 (n = 26) | 0.311 |
| Relevant personal situation | 4.3 SD = 1.91 (n = 30) | 3.6 SD = 2.19 (n = 5) | 4.4 SD = 1.87 (n = 25) | 0.341 |
| Motivational | 4.2 SD = 1.81 (n = 30) | 3.2 SD = 1.64 (n = 5) | 4.4 SD = 1.8 (n = 25) | 0.184 |
| Interesting | 4.0 SD = 1.62 (n = 30) | 3.0 SD = 1.73 (n = 5) | 4.2 SD = 1.55 (n = 25) | 0.134 |
| Believable and trustworthy | 4.8 SD = 1.40 (n = 29) | 4.6 SD = 1.95 (n = 5) | 4.8 SD = 1.31 (n = 24) | 0.439 |
| About obstacles encountered | 3.6 SD = 1.70 (n = 29) | 2.8 SD = 1.64 (n = 5) | 3.8 SD = 1.68 (n = 24) | 0.218 |
| Suitable for age and perceptions | 4.17 SD = 1.62 (n = 30) | 4.2 SD = 2.28 (n = 5) | 4.2 SD = 1.52 (n = 25) | 0.456 |
| *Average score topic* | *4.4 SD = 1.72* | *3.9 SD = 1.86* | *4.5 SD = 1.68* | *0.000 ** |

The analysis of the evaluation of the persuasive technology questionnaire shows that experienced effectiveness, user-friendliness of the tablet app, notifications, and gamification were evaluated positively, but the foodbox was evaluated negatively. See Figure 6.12. The differences in evaluation for communication styles was analysed (Table 6.8). The COM style was used for 30 participants, and the AUT style was used for 6. The COM style was evaluated significantly better compared with the AUT style (*p*-value = 0.000).

#### 6.4.4.2 Differences between tablet conditions for the experience with the technology

For the experienced effectiveness and the user-friendliness of the tablet, the differences between the tablet conditions were analysed. As the usage of the foodbox was the same in both tablet conditions, the differences for the foodbox questions were not studied. The same holds for the notification messages.

Table 6.9 shows the average scores for the questions about experienced effectiveness and user-friendliness of the tablet for both tablet conditions. It also shows the combined scores for both topics. Only for the question about the motivation to stick to the dietary advice was a significant difference found. The gamification condition gave a higher score to this question.

Statistical tests were performed for the difference between conditions for experienced effectiveness and user-friendliness of the tablet app (Table 6.9). The results show that participants from the gamification condition rated experiencing motivation significantly better ($p$-value = 0.037).

#### 6.4.4.3 Differences by personal characteristics for the experience with the rechnology

6

It is interesting to see if the evaluation of participants is influenced by personal characteristics. For this, several statistical tests were performed: the Mann–Whitney U test for the difference between men and women, and Spearman correlations for the other characteristics. For the questions about experienced effectiveness, a significant difference was found: men evaluated the experienced effectiveness higher compared with women ($p$-value = 0.016). Moreover, there was a significant positive correlation ($p$-value = 0.030) between the age of participants and the motivation to stick to the dietary advice provided by the tablet (Q1).

For the questions about the user-friendliness as well as for the foodbox and for the averages for each topic, no significant differences or correlations between personal characteristics and evaluation were found.

The evaluation of notifications seems to be more dependent on personal characteristics. Two questions, asking whether any obstacles were encountered (Q20) and whether the messages were suited to the participant's age and perception (Q21), were evaluated significantly higher by women ($p$-value = 0.028 and $p$-value = 0.008). Moreover, the average score for the messages was significantly higher for women ($p$-value = 0.001). Moreover, a significant negative correlation between the protein intake at baseline and the question about the friendly tone of messages (Q14) was found ($p$-value = 0.023). Finally, there were different significant negative correlations between education level and the evaluation of messages: whether messages were relevant (Q16—$p$-value = 0.000), were motivational (Q17—$p$-value = 0.007), were interesting to read (Q18—$p$-value = 0.002), asked about any obstacles faced (Q20—$p$-value = 0.001), or were suitable for age and perceptions (Q21—$p$-value = 0.037). Moreover, the combined evaluation was also significantly negatively correlated ($p$-value = 0.005).

**Table 6.9:** Average scores for the follow-up questions about the tablet for different conditions—scale 1–7, 1 = strongly disagree, 7 = strongly agree. * $p < 0.05$.

| | **Both Conditions** | **Normal Condition** | **Gamification Condition** | ***p*-Value** |
|---|---|---|---|---|
| Motivation stick dietary advice (Q1) | 5.3 SD = 1.60 (n = 36) | 4.8 SD = 1.63 (n = 17) | 5.7 SD = 1.49 (n = 19) | 0.037 * |
| Functionalities needed present (Q2) | 5.0 SD = 1.71 (n = 36) | 4.65 SD = 1.69 (n = 17) | 5.3 SD = 1.70 (n = 19) | 0.105 |
| Meets expectations (Q3) | 4.5 SD = 2.08 (n = 36) | 4.2 SD = 1.89 (n = 17) | 4.8 SD = 2.25 (n = 19) | 0.175 |
| Continue to use (Q4) | 3.5 SD = 1.93 (n = 35) | 3.2 SD = 1.65 (n = 16) | 3.6 SD = 2.17 (n = 19) | 0.468 |
| *Average score experienced effectiveness* | *4.6 SD = 1.94* | *4.2 SD = 1.79* | *4.8 SD = 2.04* | *0.242* |
| Easy to use (Q5) | 5.06 SD = 1.90 (n = 36) | 4.8 SD = 1.60 (n = 17) | 5.3 SD = 2.14 (n = 19) | 0.081 |
| No irregularities encountered (Q6) | 3.7 SD = 2.09 (n = 35) | 3.5 SD = 2.10 (n = 16) | 3.9 SD = 2.12 (n = 19) | 0.368 |
| Quickly way around (Q7) | 5.6 SD = 1.50 (n = 36) | 5.3 SD = 1.61 (n = 17) | 5.9 SD = 1.37 (n = 19) | 0.103 |
| Recommend to others (Q8) | 4.5 SD = 2.13 (n = 36) | 4.2 SD = 2.19 (n = 17) | 4.7 SD = 2.11 (n = 19) | 0.269 |
| Comprehensible (Q9) | 5.2 SD = 1.76 (n = 36) | 5.0 SD = 1.62 (n = 17) | 5.3 SD = 1.92 (n = 19) | 0.195 |
| Fun to use (Q10) | 5.1 SD = 2.07 (n = 36) | 4.6 SD = 2.15 (n = 17) | 5.6 SD = 1.92 (n = 19) | 0.066 |
| *Average score user-friendliness* | *4.9 SD = 1.99* | *4.6 SD = 1.94* | *5.1 SD = 2.01* | *0.241* |

Lastly, only one significant correlation was found for the gamification questions: a negative correlation between education level and motivation of the profile page (Q22—*p*-value = 0.047).

Several significant differences and correlations are found between personal characteristics and the evaluation of the experience with the persuasive technology:

- Men evaluated the experienced effectiveness significantly higher compared with women (*p*-value = 0.016)
- There was a significant positive correlation between the age of participants and the motivation to stick to the dietary advice provided by the tablet (*p*-value = 0.030).
- Different significant negative correlations were found for education level and questions about the messages: messages were relevant, were interesting to read, asked about any obstacles faced, or were suitable for age and perceptions. The overall evaluation of the messages was also significantly negatively correlated with education level.
- The motivation provided on the profile page has a significant negative correlation with education level (*p*-value = 0.047).

#### 6.4.4.4 Interaction with the tablet

It was studied how participants used the technology, which can give insights into which elements are used more often. Figure 6.13 shows how often, relative for each user, the three different input methods (fast input screen, meal composer, or via the foodbox) are used. As this is a relative graph, the extended trial due to COVID-19 was also taken into account. This shows that foodbox input is only used by a few participants. Moreover, some participants use fast input for most inputs, while other participants use the meal composer most of the time. Only in a few cases, this is a bit more balanced. It needs to be noted that this graph does not correct for repeated inputs. Repeated inputs can be due to two reasons: a participant changes an input later or a participant repeats its input (for example to check it or to redo it).

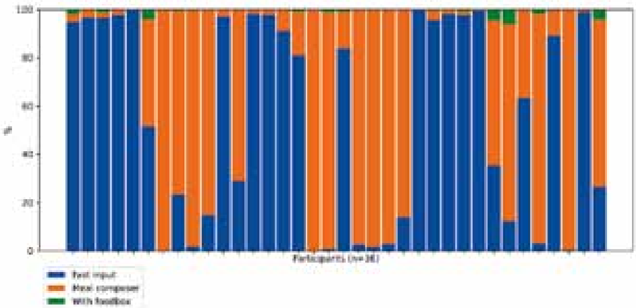


**Figure 6.13:** Relative number of registrations on the tablet with different methods of input. [10]

In general, the meal composer is used 52% of the inputs, while fast input is used for 47% of the inputs and the foodbox is only used for 1% of the inputs.

Participants can react in two ways to a notification: they can either confirm, meaning they start to enter their intake, or postpone the notification, which enables a reminder. Of the total number of notifications sent, 87% was ignored, 12% was confirmed, and 1% was postponed.

To study how the gamification components were used, different types of interaction were studied. The data were not cleaned to take into account the extension of the trial due to COVID-19. Only two participants in the gamification condition fell into this category. It was however checked whether their data were different from the other participants.

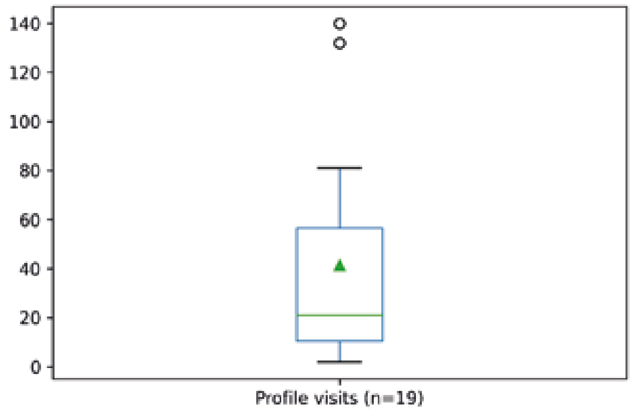



**Figure 6.14:** Boxplot of profile page visits for the gamification-condition participants. [10]

It was studied how often the participants visited their profile page. Figure 6.14 shows that there were quite some participants that only visited it less than 20 times. There were also two extremes of participants who visited it more than 100 times, which was not caused by the extension of the trial due to COVID-19. On average, participants visited their profile page 41.5 times (SD = 41.61). It can be tested whether there was a Pearson correlation between the days with inputs and the number of profile visits for participants. This correlation (0.032) was not significant (*p*-value = 0.179). This means that the number of profile visits does not correlate with the number of days that the participants use their tablets.

When a participant's intake met the protein criteria for that eating moment, a mini-game is triggered. When asked to play a game, it is also possible to skip the game. Moreover, it was found that not all games that are started are also finished. This can be due to a bug, which closes the screens when the mini-game was offered to the participant, but it can also be because the participant put the tablet aside during the game. Figure 6.15 shows that only six participants (32%) of this condition played more than 50% of the mini-games that they were allowed to play. Some participants did not finish a single game.

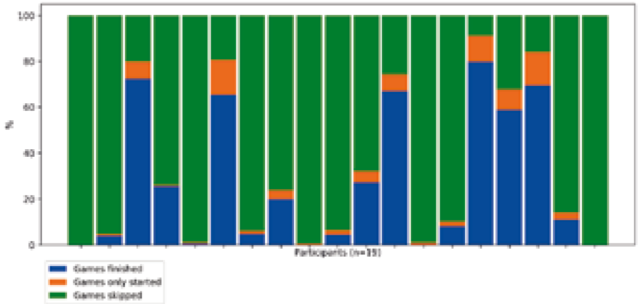



**Figure 6.15:** Relative number of mini-games played and skipped per gamification condition participant. [10]

In Figure 6.16, it is analysed whether participants have a preference for a specific mini-game. To analyse this, the number of finished games was compared with the total number of proposed games (the games finished plus the games that were skipped). The games that were started but not finished were ignored because whether this is a bug or intentional cannot be detected. For some participants, this figure shows that they have a preference for some games over other games, but this pattern is different for each participant.



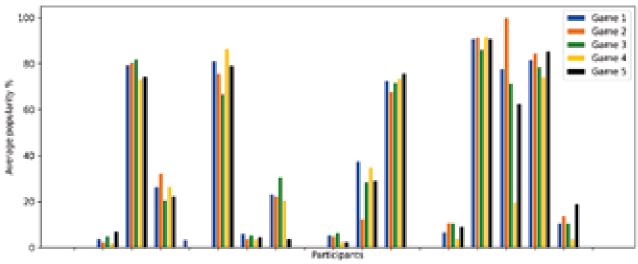



**Figure 6.16:** Popularity difference for mini-games.

Figure 6.17 shows that the popularity of all games, on average, is comparable. For this graph, all popularity scores are summed and divided by the number of gamification participants. To compare the averages a one-way ANOVA test was performed. This showed that no significant difference between the different averages could be found ($p$-value = 0.994).

As the studied aspects of interaction are either the same in both conditions or only present for the gamification condition, the differences between the two study groups for these aspects were not studied.

For the different methods of input (Figure 6.13), again, it was shown that the foodbox was not often used and that the other methods were almost equally often used. Overall, participants had one preferred way of input (fast input or meal composer) and only some used them both in a balanced manner. Most participants did not use notification but instead choose when to input their consumption when it suited them: 87% of the

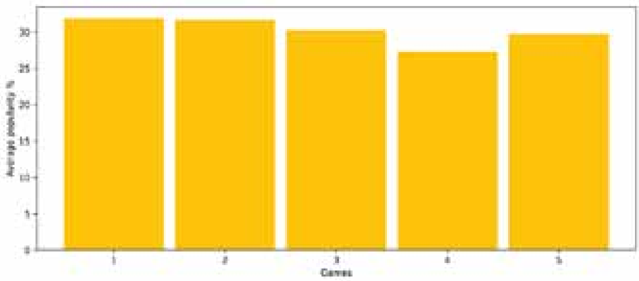


**Figure 6.17:** Average popularity of each mini-game.

notifications were ignored, 12% were confirmed, and 1% were postponed. The number of profile page visits is presented in Figure 6.14. The average number of visits is 41.5 (SD = 41.61, minimum = 2, maximum = 140). The Pearson correlation between days of use and profile visits is not significant ($p$-value = 0.179). For the gamification, some participants liked and played the mini-games, while others did not or played only some games (Figure 6.15). For the popularity of the games, only slight differences can be seen for some participants (Figure 6.16), but overall (Figure 6.17), no significant differences were found ($p$-value = 0.994).

#### 6.4.4.5 Differences by personal characteristics for interaction

To see if the way that participants interact with the tablet is dependent on personal characteristics, several statistical tests were performed. First, differences between males and females were studied using Mann–Whitney U tests. For the percentage of inputs carried out using the fast input method, a significant difference was found ($p$-value = 0.031). Male participants used a fast input on average a bit more (mean = 55.8%, SD = 42.83%) compared with female participants (mean = 55.3%, SD = 43.52%). Sex differences were not found for the number of profile visits or the percentage of mini-games finished.

Next, tests for significant Pearson correlations were performed for age, protein intake at baseline for the percentage of fast input, number of profile visits, and percentage of mini-games finished. A one-way ANOVA was used to compare the different education levels for these measures. No significant results were found.

Different statistical tests for the influence of personal characteristics on the interaction with the tablet were performed. The only significant result is that fast input is used significantly more by male users ($p$-value = 0.031).

### 6.4.5 Gamification and protein knowledge (RT5)

In the dietary advice evaluation questionnaire, one question addressed whether the participants feel that they have a good understanding of the amount of protein in different types of products (see Appendix 6.A.1). This question gives insight into whether participants in the PT group gained more insight into protein products compared with the no-PT group.

In contrast to our expectations, the mean score of the PT group (n = 36, mean = 3.9, SD = 0.93) was lower compared with the no-PT group (n = 37, mean = 4.1, SD = 0.74); this difference is significant ($p$-value = 0.030). The normal condition (n = 17, mean = 3.9, SD = 0.83) was compared with the gamification condition (n = 19, mean = 3.8, SD = 1.03), but these results turned out not significant ($p$-value = 0.391).

Another way to study whether there was a learning effect for the mini-games is to look at the scores for the mini-games. To do so, the mean per five mini-games of one type was calculated. For participants who played at least one mini-game more than five times, a graph was created, which is shown in Figure 6.18. Overall, there was no trend towards higher scores when playing more mini-games. However, when looking at participants who play the mini-games quite often, it seems that their scores often stabilised.

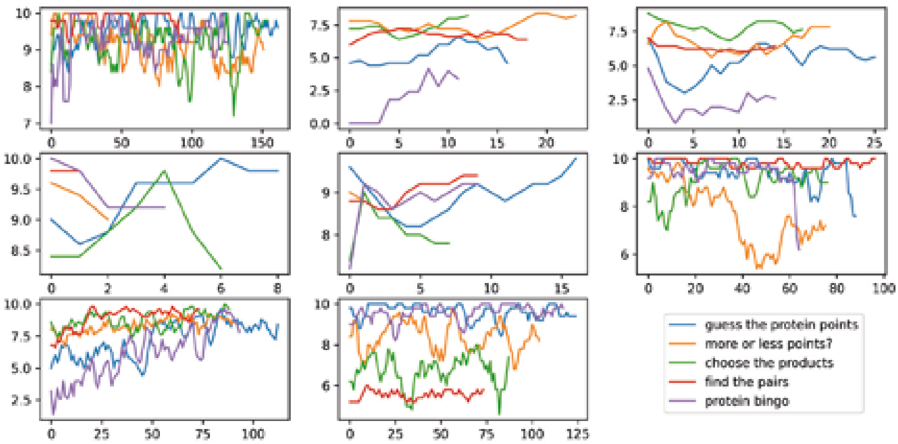



**Figure 6.18:** Average scores per five games played for gamification participants with more than five games played—each graph represents the scores of one player. The X-axis shows the number of games played, while the y-axis shows the scores (0–10).

Next, the scores in the first half of the games and the scores in the last half of the games were analysed. For this, all games were combined per participant. With a $t$-test, it was tested whether the differences between those means were significant. All participants from the gamification condition who played at least two games were included (14 participants). From these tests, it becomes clear that five participants made a significant change in their average points. For four of these participants, this is an improvement; for one, it is not, but this participant played a very low number of games compared with the other participants. It needs to be noted that the scores are not a complete representation of the knowledge of participants. It is unclear whether participants used their protein information booklet when playing the games or whether this change is because they understood the games better.

Whether there is a Pearson correlation between the number of games played and the average score of participants was tested. The coefficient was 0.689, with a significant *p*-value of 0.006. Figure 6.19 shows the scatterplot of this analysis. Again, although the correlation was significant, it cannot be concluded based on this that the games contributed to more protein knowledge, as other explanations for this finding are also possible, for the reasons explained above.

Participants in the tablet condition rated their protein knowledge significantly lower compared with participants without a tablet (*p*-value 0.030). No significant result was found between the two sub-study conditions concerning their self-reported protein knowledge (*p*-value = 0.391). When looking at the scores that participants earn in the mini-games, it is hard to observe any trends, but when a high number of games were played, the scores seemed to stabilise (Figure 6.18). Comparing the mean scores of the first and second halves of the games played per participant, significant differences can be found for 5 out of 13 participants. Four of those participants improved their scores in the second half. A significant Pearson correlation was found between games played and average game score (Figure 6.19, coefficient 0.674, *p*-value = 0.012).

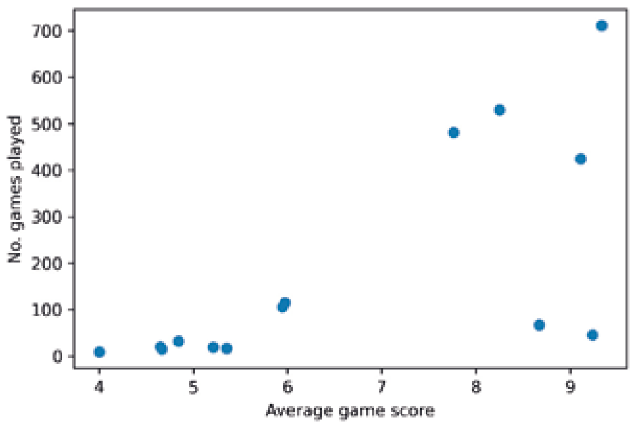



**Figure 6.19:** Scatterplot average game score and number of games played (n = 14).

### Differences by personal characteristics for protein knowledge

Differences in the rating of the question about understanding protein points caused by personal characteristics were tested. With the Mann–Whitney U test, it was tested if there were differences caused by sex, but this was not significant (*p*-value = 0.472). Next, the Spearman correlation was calculated between age, protein intake at baseline, and education level, and rating for their understanding of the questions. Again, these correlations were not significant.

Moreover, to see if personal characteristics influence the average mini-game score, a Mann–Whitney U test was used for differences between male and female participants, a Pearson correlation was used for the age and protein intake at baseline, and a one-way ANOVA was used for education level. All tests were not significant, so personal characteristics do not have a significant effect on the average game scores of participants.

No significant results were found in the statistical tests for the association between personal characteristics, and reported protein knowledge and average game scores.

## 6.5 Discussion

This chapter gives an overview of the results of the persuasive technology sub-study of the PROMISS trial. The aim was to learn more about how older adults receive a diet-tracking system as part of a diet program, in our specific case, the PROMISS trial. Therefore, the adherence of participants, the change in protein intake, the experience of the diet and the persuasive technology, and the impact of gamification on protein knowledge were studied in this chapter. If relevant, it was studied if differences between the normal and gamified conditions exist, and the possible effects of personal characteristics were studied as well.

The participants of the study were 83% adherent to using the technology during the trial period, with an average of 133 days. No differences in the adherence of participants were found between the two conditions in the study or based on personal characteristics.

Although the adherence was high, using the technology did not significantly change their protein intake compared with participants in the intervention groups without a tablet. No effect was found for the number of days that the tablet was used on the protein intake. During the study, it was noted that the tablet sometimes gave higher estimates of protein intake due to round differences with the dietitians. However, the results show that this did not have a significant negative effect on the protein intake of participants. No differences between the two conditions or caused by personal characteristics for the protein intake were found.

Participants with or without a tablet did not significantly evaluate the differently diet on a whole. The normal condition in the PT group gave the diet a significantly higher rating compared with the gamification condition. Moreover, it was found that women in the PT group have a significantly higher intention to continue to follow the dietary advice, while this difference was not found in the no-PT group.

Participants appreciated the tablet and the gamification elements, but the foodbox was negatively evaluated. For the experienced effectiveness and the user-friendliness of the tablet, differences between the two conditions were studied. Only one significant difference was found: the gamified condition reported a higher motivation to stick to the dietary advice by the tablet, compared with the normal condition. Furthermore, for all questions, it was tested whether personal characteristics influence the evaluation of the persuasive technology. Some effects of sex and age were found, but the most significant negative correlations were found for education level and the questions about the notification messages.

When studying the interaction with the app, the data showed that participants often ignore the notification message, and instead interacted with the app on their own initiative. Most participants either used the fast input or the meal composer most of the time, some participants used these methods of input in a more balanced manner. For the mini-games, a similar pattern was seen: some participants skipped all mini-games, other participants played almost all mini-games, and some participants played some mini-games and skipped some others. A significant difference was found in the percentage of inputs using the fast input between men and women (higher for men).

Finally, the learning effect of the mini-games was studied. Based on a question about the understanding of protein, this cannot be derived. From the scores of the games, it seemed that there is a learning effect for some participants. Moreover, the scores were correlated with the number of games played. However, higher scores can also be due to other aspects, such as using the information booklet. Altogether, the app was considered a feasible way to teach about protein intake. No effects of personal characteristics on the rating of the understanding or the scores in the mini-games were found.

There were some limitations to the study. The sample size of the sub-study was 36, which was limited but turned out to be sufficient for the presented analysis. When the different conditions were studied, the sample size was small (17–19). Therefore, it is unclear whether the analysis would hold with a larger sample. Moreover, it might be that the participants had some different characteristic(s) that made them volunteer to participate in the study, which could affect the results. This could not be derived from the available data.

During the study, it was sometimes necessary to fix some problems with the application. The inclusion for the gamification condition started later in the study after some bugs were fixed. The evaluation and participation of participants may have been influenced by the number of bugs they encountered during their trial period. However, no major differences were found in the average scores of the evaluation questionnaire.

For the analysis of the adherence, the 12 drop-out participants were not included. Including all of those excluded participants changed the overall adherence to 104 days and 76%. However, four participants did not have any interaction with the system. When excluding those, the average adherence was 114 days and 83%. Although the average adherence expressed in active days would have lowered, it was still more than half of the trial period (6 months) and the ratio of active days was the same as the adherence ratio of the PT participants. The results on adherence can also be influenced by the way the duration of the tablet trial was counted for each participant. The first and last days of use were used to calculate the trial duration. It could be that participants put the tablet aside for some time and started using it again right before an appointment. However, this was not found in our data set. Some participants did not use the tablet for a longer period during their trial, sometimes due to issues or holidays and sometimes without a given reason. However, they started using it again for a considerable time before ending their use. For only one participant, it was found that there were around two weeks of inactivity, then five days of activity, and then the participant stopped using the tablet.

This research does not look at the quality of the diet tracking data that participants entered into the app. It was not studied whether the data from the tablet could be used by dietitians within the trial. For example, dietitians perform phone recalls during the trial to discuss the protein intake of participants. It would be interesting to study if the data from the tablet could replace such recalls. However, another research setup would be needed to address this.

## 6.6 Conclusions

Based on the results, we conclude that using a tablet with a diet-tracking application, with persuasive communication, dedicated to the trial is a feasible way to engage participants. The target users, older adults, support the use of such persuasive technology, as shown by their evaluation of the system as well as their adherence to using it. The experienced effectiveness and the user-friendliness of the tablet were evaluated positively. The adherence was 83%, with an average of 133 days. No major differences in the evaluation of the normal condition and the gamification condition were found for the adherence or evaluation of the system or diet. Moreover, it is shown that the gamification elements were not liked/used by all participants and seems to be a personal preference. As no personal characteristics were found that might be correlated with this, it cannot be predicted which participants will like it. However, it does not seem to be a drawback for participants who do not like the gamification. For all different measures, associations with personal characteristics were studied. Although a few significant correlations were found, no patterns were found for specific groups who experience or evaluate the technology differently. Multiple significant correlations between the evaluation of the notification messages and education level were found, so this can be something to investigate in further research. Overall, it can be concluded that using diet-tracking applications in trials is well received by older adult participants and is thus a feasible way to track a diet. In future research, it would be interesting to study how such persuasive technology can be further integrated into a diet study, for example, to support the work of dietitians.



## 6.A Questionnaires

### 6.A.1 Dietary advice evaluation

This questionnaire consists of seven questions about the following measures:

- Rating dietary advice (scale 1–10);
- Usefulness dietary advice;
- Extent to which dietary advice was followed;
- Ease of sticking to the dietary advice;
- Intent to continue following dietary advice;
- Understanding of protein in products;
- Ease of finding products with similar protein amounts.

If not indicated otherwise, the measure was evaluated on a five-point Likert scale. To analyse the Likert-scale questions, the Mann–Whitney U test was used. For the rating, a $t$-test was used.

### 6.A.2 Persuasive technology questionnaire

Different topics were discussed in this questionnaire. The following topics and questions were asked (translated from Dutch):

- Effectiveness of the tablet application
  1. The tablet helps/motivates me to stick to my dietary advice.
  2. The tablet has all of the functionalities that I need.
  3. The tablet meets my expectation.
  4. If I continue to follow dietary advice, I would like to keep on using the tablet.
- User-friendliness of the tablet
  5. The tablet was easy to use.
  6. I did not encounter any irregularities when using the tablet.
  7. I quickly knew my way around the tablet.
  8. I would recommend the tablet to someone else.
  9. The tablet was comprehensible.
  10. It was fun to use the tablet to follow my dietary advice.
- Foodbox
  11. If I stick to my dietary advice, I would like to continue using the foodbox.
  12. The foodbox is easy to use.
  13. It was fun to use the foodbox when following my dietary advice.
- Notifications
  14. The messages have a friendly tone.
  15. The messages are compelling.
  16. The messages are relevant for my personal situation.
  17. The messages are motivational.
  18. The messages are interesting to read.
  19. The messages are believable and trustworthy.
  20. The messages are about obstacles that I encounter.
  21. The messages suit my age and perceptions.
- Gamification
  22. I was motivated by my profile page.
  23. I was motivated by achieving an achievement.
  24. I found playing games motivating.

25. I liked playing the games.

26. I found the games informative.

All questions for this questionnaire could be rated on a 7-point Likert scale (1 = strongly disagree, 4 = neutral, 7 = strongly agree). Question 15 is a negatively framed question, and the scores are therefore reversed in the analyses. A Mann–Whitney U test was used to compare the mean scores of the different groups, if applicable.

## Bibliography

[1] Al-Razgan, M.S., Al-Khalifa, H.S., Al-Shahrani, M.D., AlAjmi, H.H.: Touch-based mobile phone interface guidelines and design recommendations for elderly people: A survey of the literature. In: Int. Conf. Neural Inform. Processing. Springer (2012)

[2] Carter, M.C., Burley, V.J., Nykjaer, C., Cade, J.E.: Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. Journal of medical Internet research **15**(4) (2013)

[3] Cordeiro, F., Epstein, D.A., Thomaz, E., Bales, E., Jagannathan, A.K., Abowd, G.D., Fogarty, J.: Barriers and negative nudges: Exploring challenges in food journaling. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. pp. 1159–1162. ACM (2015)

[4] Coughlin, S.S., Whitehead, M., Sheats, J.Q., Mastromonico, J., Hardy, D., Smith, S.A.: Smartphone applications for promoting healthy diet and nutrition: a literature review. Jacobs journal of food and nutrition **2**(3), 021 (2015)

[5] Johnson, D., Deterding, S., Kuhn, K.A., Staneva, A., Stoyanov, S., Hides, L.: Gamification for health and wellbeing: A systematic review of the literature. Internet interventions **6**, 89–106 (2016)

[6] Johnson, R., Kent, S.: Designing universal access: web-applications for the elderly and disabled. Cognition, Technology & Work **9**(4), 209–218 (2007)

[7] Koivisto, J., Hamari, J.: Demographic differences in perceived benefits from gamification. Computers in Human Behavior **35**, 179–188 (2014)

[8] van der Lubbe, L.M., Klein, M.C.A.: Designing a system with persuasive communication to improve diet compliance for elderly users. In: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. p. 234–241. PervasiveHealth'19, Association for Computing Machinery, New York, NY, USA (2019)

[9] van der Lubbe, L.M., Klein, M.C.A.: Integrating gamification into a system to improve diet compliance for elderly users. In: Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good. p. 150–155. GoodTechs '20, Association for Computing Machinery, New York, NY, USA (2020)

[10] van der Lubbe, L., Klein, M., Visser, M., Wijnhoven, H., Reinders, I.: Experiences with using persuasive technology in a diet trial for older adults. In: The 14th PErvasive Technologies Related to Assistive Environments Conference. p. 244–251. PETRA 2021, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3453892.3458686

[11] Reinders, I., Visser, M., Jyväkorpi, S.K., Niskanen, R.T., Bosmans, J.E., Jornada Ben, Â., Brouwer, I.A., Kuijper, L.D., Olthof, M.R., Pitkälä, K.H., et al.: The cost effectiveness of personalized dietary advice to increase protein intake in older adults with lower habitual protein intake: a randomized controlled trial. European journal of nutrition pp. 1–16 (2021)

[12] Reinders, I., Wijnhoven, H.A.H., Jyväkorpi, S.K., Suominen, M.H., Niskanen, R., Bosmans, J.E., Brouwer, I.A., Fluitman, K.S., Klein, M.C.A., Kuijper, L.D., van der Lubbe, L.M., Olthof, M.R., Pitkälä, K.H., Vijlbrief, R., Visser, M.: Effectiveness and cost-effectiveness of personalised dietary advice aiming at increasing protein intake on physical functioning in community-dwelling older adults with lower habitual protein intake. BMJ Open **10**(11) (2020)

[13] Seaborn, K., Fels, D.I.: Gamification in theory and action: A survey. International Journal of human-computer studies **74**, 14–31 (2015)

[14] Stanmore, E.K., Mavroeidi, A., de Jong, L.D., Skelton, D.A., Sutton, C.J., Benedetto, V., Munford, L.A., Meekes, W., Bell, V., Todd, C.: The effectiveness and cost-effectiveness of strength and balance exergames to reduce falls risk for people aged 55 years and older in uk assisted living facilities: a multi-centre, cluster randomised controlled trial. BMC medicine **17**(1), 49 (2019)

[15] Van Genugten, L., Dusseldorp, E., Webb, T.L., Van Empelen, P.: Which combinations of techniques and modes of delivery in internet-based interventions effectively change health behavior? a meta-analysis. Journal of medical Internet research **18**(6), e155 (2016)

[16] Webb, T., Joseph, J., Yardley, L., Michie, S.: Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. Journal of medical Internet research **12**(1) (2010)

# Chapter 7

# Improving the recommendations of meals in the PROMISS application

This chapter is reviewed and accepted for Goodtechs 2021 as: Spooren, D., **van der Lubbe, L.M.** Improving the Recommendations of Meals in the PROMISS Application

**This chapter contains:**



## Abstract

The PROMISS application is specifically built to let older adults keep track of their diet and protein intake. To improve the user-experience of this application, we study how machine learning algorithms can be used to recommend meals and products based on

historical data. An intelligent workflow is designed which combines five different algorithms that recommend suitable meals and products. These algorithms are trained and tested using data from a previous user study with the PROMISS application. The change in user-experience is measured by the numbers of clicks needed to enter a meal in the application. Two different variants of the new application, namely, one using only the two new recommended meals and the other using both the two new recommended meals plus the old recommended meal, are compared with the old application. It was found that both new applications reduce the number of clicks and thus increase the user-experience of the application.

## 7.1 Introduction

Over the last few years, the usage of technology in the lives of older adults has increased significantly. An example of this is PARO: the famous seal that is used in elderly homes as a companion robot and a therapeutic tool. Research showed that the older adults were willing to interact with PARO and that this interaction improved their physical activity [12].

Technology can also be used in order to stimulate a healthy lifestyle, for example meal planning systems. Such an application has been created for the diet trial of the PROMISS project (Chapter 4, [17]). The PROMISS diet trial aims on increasing the protein intake of older adults with a relatively low protein intake [19]. One of the risks of a low protein intake is a rapid loss of muscle mass [10].

The PROMISS application is meant for daily usage during the diet trial. For each user, the application is personalized with the help of their personal diet plan created by the dietitians of the PROMISS project. Using a progress bar, the total amount of protein consumed by the user is visualized throughout the day, helping the user complete their daily protein intake.

The PROMISS application that was used in the diet trail included meal recommendations based on the user's diet plan and previous input (Chapter 4, [17]). In this project the possibility of improving the user-experience by reducing the number of necessary clicks to enter a meal is studied. This can be achieved by creating an intelligent workflow using machine learning algorithms that personalize meal and product recommendations on regularities in historical data of users. Furthermore, it has been shown that using computer tailored information personalized on the user, is more effective in promoting a nutritious lifestyle, than general non-computer tailored information [11]. Thus, recommending meals in a more personalized way improves the user-experience not only by increasing the efficiency but also by making the recommendations more personalized.

First, background information about the PROMISS application and meal recommendation is discussed in Section 7.2. Second, the research methodology, including a description of the data, is described in Section 7.3. Third, the different algorithms are described in Section 7.4, together with an overview on how they work together. In Section 7.5 it is discussed how the training data has been determined. Finally, the conclusion and discussion can be found in Section 7.6.

## 7.2 Background

First, more details on the PROMISS application are discussed. Furthermore, an overview of related work on meal and product recommendation is given.

### 7.2.1 PROMISS application

As mentioned in the introduction, the PROMISS application is a system to improve diet compliance for elderly users (Chapter 4, [17]). Protein points were used as a way to represent the protein value of products and meals. The users of the application could keep track of their diet by means of their protein points. The goal was to stimulate users to eat enough protein each day by providing them with a progress bar of the protein points they have gotten on a specific day or moment.

Before the diet trial, the eating habits of participants were monitored. Based on this information a diet plan, including the personal protein need, for each user was composed by a dietitian. For most participants, there were six eating moments per day: three main meals and three snacks. The diet plan consisted of one meal for each eating moment of the day, and it was the same for each day. The users each received a tablet with their personalized application. They were instructed to fill in meals they have eaten for each eating moment for a specific period of time. The original application contained three ways of entering meals:

- The user can enter a meal via the meal composer. They can replace, remove or add products from the recommended meal. When the user has switched a product for three times or more for the same product, the meal plan for that eating moment is adjusted by replacing the product.
- The user can directly enter the number of protein points eaten, without filling in the products of the meal.
- The user can use the additionally provided foodbox to register intake of specific products.

This system also had disadvantages. Firstly, each day the recommended meal was, in essence, the same. Secondly, when the user chose to deviate from the recommended meal, (s)he had to enter this meal manually. This can be a time depending task.

### 7.2.2 Meal and product recommendation

Technology is growing each day and is playing an increasingly larger role in our lives. Whether needed for work, sharing pictures on social media or downloading useful applications to make life easier, nearly everyone owns a mobile device [2]. Another thing that keeps growing is the problem of obesity and poor health. Being obese causes the death of over four million people each year [4]. However, it has been shown that obesity and health related problems can be prevented or even reversed in some cases through good nutrition [18]. Because of the growing use of technology and health problems due to obesity, there are quite some food recommender tools on the market which all try to stimulate a healthy lifestyle.

Elsweiler and Harvey presented an approach to integrating nutrition in a recommender system by grouping items, which together present a balanced meal, rather than recommending individual items [13]. During this experiment Elsweiler and Harvey gathered a taste profile of the participants, using data including recipes and nutritional properties, which users could rate. One potential pitfall was that the users did not rate enough breakfast recipes, and it has not been further researched how to improve the recommendation for breakfast meals.

Furthermore, the approach of Elsweiler and Harvey did not contain any research on the usage of the application by older adults [13]. The system 'Nutrition for Elder care' (NutElCare) by Espin, on the other hand, has been especially designed for older adults [14]. In the last few years, the usage of technologies among older adults has increased substantially. Which caused a higher willingness of using these kinds of technologies, such as food recommender systems, in their daily lives [14]. NutElCare uses knowledge-based techniques and requirements of the user to generate a recommendation of items. Furthermore, the user can rate meals which are used to calculate similarity scores between meals.

Many food recommender tools rely on the input from users. See, for example, work by Freyne and Berkovsky, whose recommender system use ratings on both recipes and food items [15]. Or the food recommender system from Professor Aberg, whose system uses collaborative filtering to predict a user's taste opinion on a recipe that the user has not yet rated based on other ratings [8]. Disadvantages of these systems are the lack of willingness of rating meals and the lack of rated meals for specific categories. No examples are found of systems that rely purely on historical data on eating habits of the users.

## 7.3 Method

This research studies whether personalized recommendations improves the user-experience of a diet tracking app. To do so, data from the PROMISS application used in the PROMISS diet trial is used. This data is described in the next section. The evaluation of the designed algorithms is explained in Section 7.3.2.

### 7.3.1 Data usage

The data that is used has been gathered from the users that participated in the PROMISS diet trial [16]. The data has been anonymised in order to maintain the anonymity of the participants. When designing the five algorithms, the data of the protein products, the activity logging of the user and the day totals of the protein points is used. From the activity logging data, especially the data where the user fills in a meal using the meal composer are important for this research. This data contains all information (e.g. product and portion size) on the meals that are entered in the meal composer.

Furthermore, not all data of all participants is used. In the tablet study of the PROMISS diet trial, 36 participants were considered active users [16]. The data varies from 27 days as the least amount of data and 240 days as the most amount of data. However, some people did not use the meal composer for the majority of the time. They entered the number of protein points without the products their meal consisted of. Therefore, the percentage of meals entered using the meal composer has been calculated for each participant. It has been decided to use a threshold of 70% and only use data of users above that threshold. Further in this research, the 11 remaining participants are referred to as active users.

Subsequently, the period of training data has to be established. There needs to be a balance between having enough data which can lead to logical recommendations and containing the satisfaction of the user. During the training period there cannot be personalized recommendations based on historical data and thus this also needs to be taken into consideration. Until enough data is present, users are considered 'new users'. In the results section is explained how the training period has been established.

### 7.3.2 Evaluation of the task completion time

When evaluating the task completion time, the number of clicks when entering a meal using the old application is compared to the number of clicks using the new application. As mentioned, this new application is improved through the use of five different algorithms which are explained in Section 7.4.




**Figure 7.1:** Screenshot of the meal composer

In Figure 7.1, a screenshot of the meal composer in the old application is shown. On the left all products in the recommended meal are shown. When clicking on the blue button below the meal ('add a new product'), the categories of products appear on the right. By clicking on a category, the top 5 most used products for that user appear followed by all products in alphabetical order. By clicking on a product, this product will be added to the meal and the right part of the screen is emptied again. Moreover, if a user wants to replace a product in the meal, (s)he can click on that product and all products from the same category will appear, with the right amount so that it contains the same protein value as the original product. In Algorithm 1 it can be seen how the number of clicks is acquired using the old application.

For evaluating the number of clicks for the new application, the test data is used to acquire the chosen meal. From the logging data of the user, it can be determined what products have been added or deleted to the recommended meal, this way the chosen meal can be acquired. Furthermore, the two personalized recommended meals obtained using the newly created algorithms are used. For the evaluation, the meal that resulted in the least number of clicks is used. A difference between the old and the new application can be seen in Algorithm 2: adding a product can also lead to two clicks when a product is chosen from the predicted product list which is acquired using a machine learning algorithm named association rule learning. When comparing the number of clicks between the old application and the new application, a two-sided test for the null hypothesis that two independent samples have identical average expected values is conducted. The test assumes that the samples have identical variances.

## 7.4 Implementation

Five different algorithms have been designed and composed together such that it creates an intelligent workflow. The code of the algorithms and the evaluation of the algorithms can be found on GitHub [1]. The algorithms are implemented in Python 3.7.6 [5]. In this section, an overview of the general workflow and a description of the five algorithms is given.

### 7.4.1 Overview and general flow

The basic working of the five algorithms is as follows:

- The **preset algorithm** computes the 10 most used presets from all active users. In this context, a preset is a combination of categories which often occur together. From these categories the algorithm creates a meal by looking at the most used products within these categories.

**Algorithm 1** Pseudocode number of clicks old application

```
Initialize a recommended meal for user
if user replace/add product then
    if product can be replaced with product from same category then
        Clicks += 2
    else
        Clicks += 3
if user removes product then
    Clicks += 1
```

**Algorithm 2** Pseudocode number of clicks new application

```
Initialize recommended meals for user
if user replace/add product then
    if product can be replaced with product from same category then
        Clicks += 2
    if product is in advised product list then
        Clicks += 2
    else
        Clicks += 3
if user removes product then
    Clicks += 1
```

- The **protein points algorithm** is only used for meals filled in for dinner and the evening snack. The algorithm looks at how much points the user has still left for that day and recommends a meal within a range of these points.
- The **core + addition algorithm** considers products which are often used together in one meal. From this, one core of two products is chosen. Furthermore, the additions are products that appear in the same meal as the core. Combining the core and some additions, one meal is created.
- The **preference algorithm** is used to recognize people with a vegetarian or a pescatarian diet. Where vegetarians exclude both meat and fish from their diet, pescotarians do eat fish but not meat. This knowledge can be used to match their recommended meals with their preference.
- The **association rule learning algorithm** is used when the user decides to add or replace a product. Machine learning is used to discover relationships between the products from the current composed meal and all products in the database. These relationships can be used to predict the next product.



Figure 7.2 shows how the five algorithms are combined into one intelligent workflow. The colors which appear in the flowchart each advocate a different algorithm. The preset and protein algorithm have been combined together, which later is explained in detail.

The first step of the flowchart is to enter a username to decide whether the user is an old or a new user. This is a crucial step, because the data used in the algorithms is different for new and old users. New users do not have historical data which can be used by the algorithms. In total there are two recommended meals. First, for old users, the preference algorithm is used to acquire the diet preference of the user. Then the two meals are deducted using a combination of the preset/protein and the core + addition algorithm. Both algorithms take into account the diet preference. The preference algorithm cannot be used for new users, hence the question whether the user's diet is vegetarian or pescatarian is asked. The two meals for new users are deducted using only the preset/protein algorithm and taking into account the diet preference.

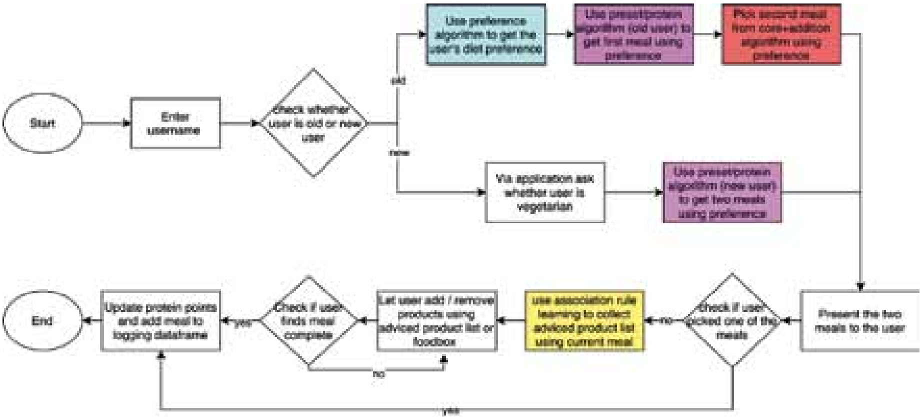



**Figure 7.2:** Flowchart of general workflow

Subsequently, the two meals are presented to the user. The user can add, replace or remove products from the recommended meal. In order to speed up this process, products are predicted using the association rule learning algorithm. This algorithm is used to acquire relationships between the products present in the meal and all the products in the database of the application. Based on these relationships a product is recommended to the user.

### 7.4.2 Preset / Protein algorithm

The preset/protein algorithm is used to recommend the first meal to old users and to recommend both meals to new users. However, the working of the algorithm is slightly different for old and new users. Figure 7.3 shows the workflow of this algorithm. The first step in the flowchart is to retrieve the eating moment which the user wants to enter a meal for, this is due to the fact that for the eating moment 'Avondeten' (dinner) or 'Tussendoor avond' (evening snack) the protein algorithm is used and in other cases the preset algorithm is used.

Assume the flow for the protein algorithm is followed, both the data of the user's day totals of protein points and the protein need is used in order to gather the amount of protein left. If this amount needs to be split between the two eating moments, then 'Avondeten' uses 75% of this amount and the 'Tussendoor avond' 25%. Afterwards the check whether a user is old or new occurs, because if the user is old, their own meals are used in order to gather all meals for the specific eating moment. If the user does not have this data, meaning the user is a new user, the meals of all active users is used.

The meals for the specific eating moment are divided into clusters using K-means clustering. This machine learning algorithm clusters data by trying to separate samples in groups of equal variances (using [7]). It requires the number of clusters to be specified, therefore a range between 2 and 40 has been chosen and for each number of clusters k-means is used. The best silhouette score is commonly used as a factor to determine the appropriate number of clusters. For each cluster number the silhouette coefficient is calculated and plotted (using [3]). Using this plot, the highest silhouette coefficient is determined and therefore the appropriate number of clusters. Subsequently, based on the amount of protein left, it is decided which cluster is closest to this amount. The meals from this cluster are then examined whether they match the user's diet preference, and the occurrences of the meals are gathered. These occurrences are used as probability in order to choose a recommended meal using a random choice method [6].

For the other eating moments, the flow from the preset algorithm is followed. Presets are categories of products which often occurs together in the eating pattern of each user. For a Dutch person this could be the standard AVG (Aardappelen - starch, Vlees - meat but also includes fish/veggie, Groenten - greens) meal. The algorithm transforms each meal from all users for a specific eating moment into a list of the categories of the products. The 10 most used presets are selected. The approach on how to retrieve the recommended meal is different for old and new users.

If the user is old, the distribution of the presets in their own meals is calculated. The distributions are used as probabilities in order to choose a preset using the random choice

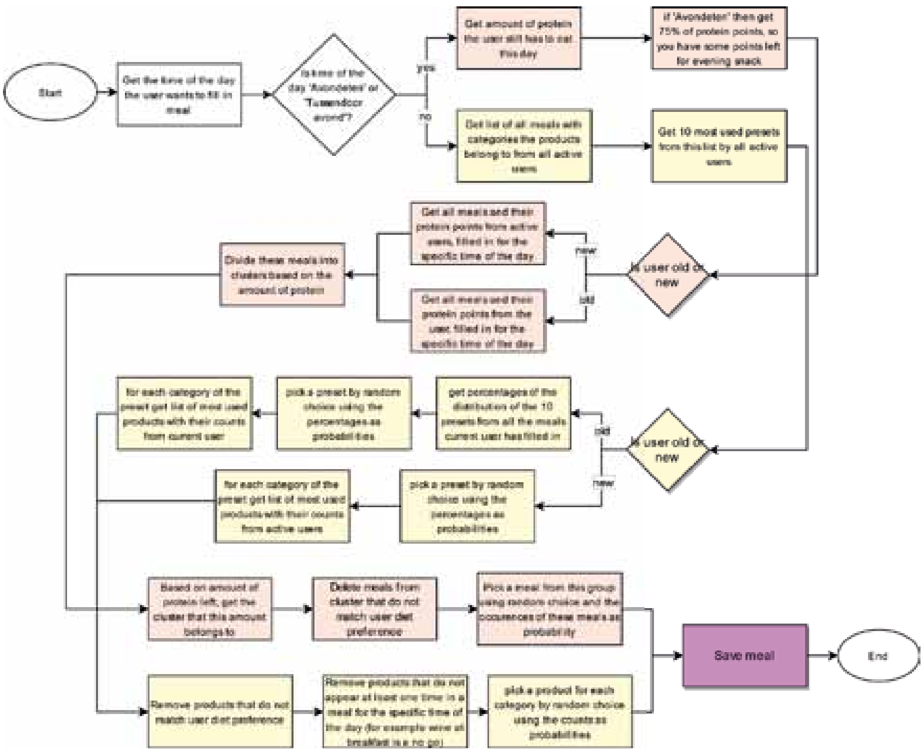



**Figure 7.3:** Flowchart of preset/protein algorithm

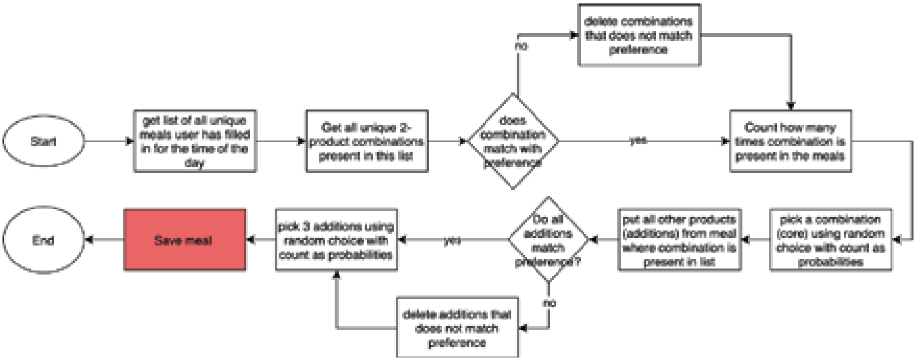



**Figure 7.4:** Flowchart of core+addition algorithm

method. For each category of the chosen preset, the most used products are gathered from all eating moments. Next, with a similar method a product is chosen for each category. The products that do not match the users diet preference are deleted. The products that do not appear at least one time in the specific eating moment are also deleted. Using the occurrences of the most used products in the meals from the user, one product for each category in the preset is chosen using the random choice method again. By adding all products together, the recommended meal is assembled.

If the user is new, so the user has not finished their training period yet, the distribution of the presets in all the meals of all users is used in order to choose a preset using the random choice method. Subsequently, the most used products for each category are gathered using all active users. From this moment, the algorithm continues the same as before. This way, users without historical data also receive a recommended meal.

### 7.4.3 Core + Addition algorithm

The core + addition algorithm is used to recommend the second meal to old users. The preset algorithm can be used for new users, using data from all active users, however the core + addition algorithm is not modelled to be able to do this. Therefore, this algorithm is only used for old users.

According to the flowchart in Figure 7.4 the first step in this algorithm is to collect all unique meals the user has filled in for a specific eating moment. From the unique meals, all unique two-product combinations are gathered. The combinations that do not match the user's diet preference are deleted. For each unique two-product combination, the occurrences in all meals from the specific eating moment from the user is collected. Using the occurrences as probabilities, one core is chosen using the random choice method. Subsequently, the algorithm goes through each meal from the specific eating moment where the core is present and gets the occurrences of all additions (consisting of one product) from these meals. The additions that do not match the user's preference are deleted. At last, three additions are collected using the random choice method and their occurrences as probability. The three additions and the core together form the recommended meal.

### 7.4.4 Preference algorithm

The preference algorithm is important for both the preset/protein algorithm and the core + addition algorithm. The algorithm predicts the diet preference of the user and therefore prevents that recommended meals for people who eat vegetarian or pescatarian contain meat or fish products. This algorithm can only be used for people who have historical data, namely the old users.

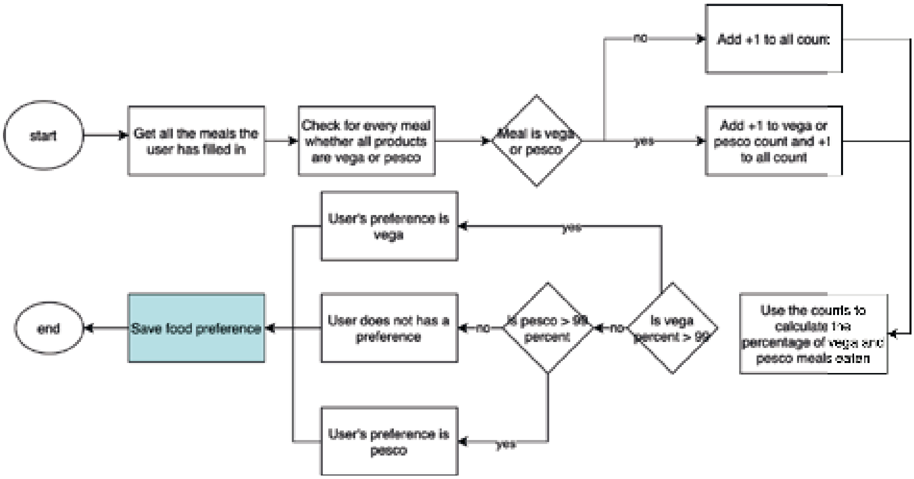



**Figure 7.5:** Flowchart of preference algorithm

First, as is shown in Figure 7.5, the algorithm collects all the meals the user has filled in and checks for every meal whether all products in the meal are vegetarian or pescatarian. Beforehand it has been decided for every product in the protein products database whether a product is vegetarian, pescatarian or neither. If the percentage of vegetarian meals is above 99%, then the user's preference is vegetarian, this is the same for a pescatarian preference. The threshold of 99% is chosen because it would be a pitfall if the threshold is not high enough and thereby will wrongly delete meat or fish products from a user's recommended meal who mainly but not entirely eats pescatarian or vegetarian. However, 100% is not used, as sometimes participants might eat something that does not match their diet preference, but still have this as a preference.

### 7.4.5 Association rule learning algorithm

The application is not only improved on the part of recommending meals, but also when the user chooses to alter the recommended meal by adding products. As can be seen in Algorithm 1, it takes the user three clicks for adding a product which is not from the same category as the product the user wants to replace. In order to reduce this to two clicks, the association rule learning algorithm is used. In Algorithm 2 it can be seen that when a user wants to add a product which is in the advised product list, it takes the user only two clicks.

There are different software implementations of the association rule algorithms. This research uses the Apriori association rule algorithm created by Christian Borgelt [9]. It proceeds by identifying the frequent individual items in a database, hence it is chosen to use this algorithm in order to improve the addition of products. The Apriori implementation uses transactional data and generates frequent item sets from within this data in order to create association rules from these item sets [20]. An antecedent is an item found within the data, in this case a product or combination of products, a consequent is an item found in combination with the antecedent. In the algorithm where Apriori has been implemented, a list is created including all consequents with a minimum support of 0.05. The support is then used as probability and using random choice two consequents are collected. For at most 5 randomly chosen products present in the current meal, their 5 consequents are collected and presented to the user, which can help them choose products for their meal.

## 7.5 Results

As mentioned in Section 7.3, the available data has been split into training and test data. This section discusses how the training period has been established. Furthermore, the number of clicks of both the old application and the newly created algorithms are presented.

### 7.5.1 Training data duration results

As previously has been discussed, a balance between having enough data which can lead to logical recommendations and containing the satisfaction of the user needs to be taken into consideration when establishing the training period. In this research the preset algorithm has been used in order to find this balance. The data of all users has been used to calculate the presets from week 0 until week 35. Week 35 being the number of days all users have used as a maximum. For each week the presets were compared to the final preset, which uses all the data from all users. The percentage of matching presets has been calculated and visualized in the graphs in Figure 7.6. It has been calculated

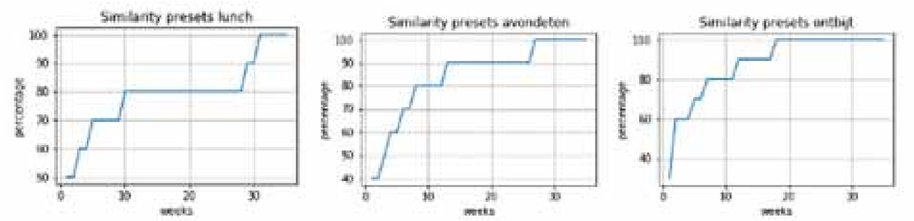



**Figure 7.6:** Similarity scores for presets ’Ontbijt’, ’Lunch’ and ’Avondeten’

**Table 7.1:** Comparison number of clicks old application and both new applications.

| | Breakfast (n=81) | Morning snack (n=61) | Lunch (n=41) | Afternoon snack (n=13) | Dinner (n=43) | Evening snack (n=12) | All meals average |
|---|---|---|---|---|---|---|---|
| **Old** | 4.8 (SD=4.32) | 6.7 (SD=5.06) | 10.1 (SD =6.81) | 5.4 (SD=3.11) | 9.62 (SD=4.55) | 3.66 (SD=3.60) | 6.71 (SD=4.57) |
| **New (2)** | 2.8 (SD=3.47) | 4.1 (SD=5.56) | 7.88 (SD=5.56) | 4.8 (SD=3.26) | 7.15 (SD=4.18) | 4.08 (SD=2.33) | 5.135 (SD=4.06) |
| **New (3)** | 2.16 (SD=2.77) | 3.72 (SD=4.94) | 7.2 (SD=5.54) | 4.6 (SD=3.36) | 6.33 (SD=3.86) | 2.9 (SD=1.76) | 4.48 (SD=3.71) |

for three eating moments, for ‘Ontbijt’ (breakfast), ‘Lunch’ and ‘Avondeten’ (dinner). This shows that for each eating moment 80% is quickly reached, afterwards the increment goes slower until it reaches 100%. For both lunch and dinner, the 100% mark takes a lot of time to reach. It makes sense in order to maintain the satisfaction of the user, to not choose 100% as a threshold. Taking lunch into consideration it is best to continue to 80%. It can be concluded that 80% is the right percentage, according to the graphs each eating moment quickly increases until 80%. Therefore, it has been decided to use 80% as a threshold, which means the training period is 10 weeks since the presets for lunch reaches this threshold at 10 weeks. Because of this, not all data can be used to evaluate the number of clicks, since some participants did not fill in meals for the whole training period. Moreover, some participants are excluded as bugs in the application makes their data unreliable. Thus, the remaining data that is used in for this study are from 7 users.



### 7.5.2 Results on the number of clicks

Algorithm 1 and Algorithm 2 show how the number of clicks can be calculated for the old and new application. In this section the results of the number of clicks are shown. For the new application, two different approaches have been used. First, the new application including the two meals resulting from the preset/protein algorithm and/or the core + addition algorithm. Secondly, a new application including three meals is evaluated. The third meal is the meal recommended in the old application and the two meals are the same as in the other new variant.

In Table 7.1 the mean average and the standard deviation of the number of clicks are shown. In the first row, the number of evaluated meals is presented. Looking at this table, it can be concluded that for each eating moment, the new application results into a lower number of clicks. Except for the eating moment ‘Tussendoor avond’: only the new application using three meals is lower than the number of clicks of the old application. Furthermore, the number of average clicks (M = 4.48, SD = 4.06) is even lower using the application using three meals than the application using two meals.

In Table 7.2 the results from the t-test comparing the new and old applications are shown The p-values which are significant ($p<0.05$), are indicated with an asterisk. The majority of the p-values of the new application using three meals are statistically significant. For example, the participants who filled in a meal for ‘Ontbijt’ (breakfast) and used the new

**Table 7.2:** Comparison p-value and t-test old application and both new applications.

| | | **Breakfast (n=81)** | **Morning snack (n=61)** | **Lunch (n=41)** | **Afternoon snack (n=13)** | **Dinner (n=43)** | **Evening snack (n=12)** | **All meals average** |
|---|---|---|---|---|---|---|---|---|
| **New (2) / Old** | p-value | 0.004* | 0.026* | 0.126 | 0.627 | 0.093 | 0.74 | 0.269 |
| | t-test | 2.904 | 2.247 | 1.545 | 0.492 | 1.698 | -0.336 | 1.425 |
| **New (3) / Old** | p-value | $8.715^{-05}$* | 0.004* | 0.048* | 0.475 | 0.005* | 0.325 | 0.142 |
| | t-test | 4.028 | 2.969 | 2.01 | 0.726 | 2.87 | 1.007 | 2.298 |

application using three meals (M = 2.16, SD = 2.77) compared to the participants who used the old application (M = 4.8, SD = 4.32) demonstrated significantly better scores (p-value = 0.004).

## 7.6 Conclusion & discussion

To improve the user-experience for a diet tracking app, this research aimed to reduce the number of needed clicks to enter a meal in the application. As a use case we used the application that is used in the PROMISS diet trial and the data that is gathered during that study. The results show that it is possible to reduce the number of clicks by improving the recommendation of meals/products. For most of the meals, this reduction is significant. The application using three recommended meals performs even better than the application using two recommended meals. Logically this can be explained by the fact that because it includes the same meal as the old app, the number of clicks can never be higher.

However another explanation of why the application using three meals performs better is the case that the chance of a recommended meal which matches the users meal, increases when the number of recommended meals rises. It could be that adding more meals even reduces the number of clicks further because of this chance. However, displaying this in a user-friendly way is challenging due to the limited screen size of a tablet or smartphone. With the previously studied version of the application [16] it was found that the elderly users were successfully able to use the application. With only a limited change to the lay-out of the meal composer, to include three menus instead of one, we do not think that this will change. However, it is important that, before using it in a live setting, we test the new lay-out with users from the target group.

In this research historical data from a previously performed study was used. However, using live data might give different results, for example for the association rule learning algorithm. For example, the order of deleted or added products could not be taken into consideration in this research, but this could lead to a different list of advised products, which could lead to a higher or lower number of clicks. Furthermore, it could be interesting to take the time of scrolling for products into consideration. This is also an important factor for the time spend on the task of entering a meal, and thus could further effect the user-experience.

Moreover, when the new application will gain users, the number of active user's needs to be set to a specific number of users which represent the whole population. When dealing with new users, both the protein and the preset algorithm uses data from active users in order to create a recommended meal. When the number of active user's rises, this could take too much time. Further research could investigate an appropriate number of users which represent the whole population of users.

This research shows that an intelligent workflow for recommending meals and products within a diet tracking app reduces the number of clicks needed to enter a meal. Although other measures such as time needed for scrolling could not be taken into account due to the lack of a live experiment, this research provides a first step in making a more intelligent diet tracking system that can be used in interventions or trials.

## Bibliography

[1] GitHub Dewi Spooren - food-recommendation, https://github.com/dewispooren/food-recommendation, assessed on: 1-7-2021

[2] How many smartphones are in the world?, https://www.bankmycell.com/blog/how-many-phones-are-in-the-world, assessed on 21-06-2021

[3] Matplotlib: Visualization with Python, https://matplotlib.org/stable/index.html, assessed on: 1-7-2021

[4] Obesity, https://www.who.int/health-topics/obesity#tab=tab_1, assessed on 21-06-2021

[5] Python 3.7.6, https://www.python.org/downloads/release/python-376/, assessed on: 1-7-2021

[6] Python library Random, https://docs.python.org/3/library/random.html, assessed on: 1-7-2021

[7] Scikit Clustering, https://scikit-learn.org/stable/modules/clustering.html#k-means, assessed on: 1-7-2021

[8] Aberg, J.: Dealing with malnutrition: A meal planning system for elderly. In: AAAI spring symposium: argumentation for consumers of healthcare. pp. 1–7 (2006)

[9] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A.I., et al.: Fast discovery of association rules. Advances in knowledge discovery and data mining **12**(1), 307–328 (1996)

[10] Baum, J.I., Wolfe, R.R.: The link between dietary protein intake, skeletal muscle function, and health in older adults. Clinical Nutrition and Aging (2017). https://doi.org/10.3390/healthcare3030529

[11] Brug, J., Oenema, A., Campbell, M.: Past, present, and future of computer-tailored nutrition education. The American journal of clinical nutrition **77**(4) (2003). https://doi.org/10.1093/ajcn/77.4.1028s

[12] Chang, W.L., Šabanovic, S., Huber, L.: Use of seal-like robot paro in sensory group therapy for older adults with dementia. In: 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI). pp. 101–102. IEEE (2013). https://doi.org/10.1109/hri.2013.6483521

[13] Elsweiler, D., Harvey, M.: Towards automatic meal plan recommendations for balanced nutrition. In: Proceedings of the 9th ACM Conference on Recommender Systems. pp. 313–316 (2015)

[14] Espín, V., Hurtado, M.V., Noguera, M.: Nutrition for elder care: a nutritional semantic recommender system for the elderly. Expert Systems **33**(2), 201–210 (2016). https://doi.org/10.1111/exsy.12143

[15] Freyne, J., Berkovsky, S.: Intelligent food planning: personalized recipe recommendation. In: Proceedings of the 15th international conference on Intelligent user interfaces. pp. 321–324 (2010). https://doi.org/10.1145/1719970.1720021

[16] van der Lubbe, L.M., Klein, M.C.A., Visser, M., Wijnhoven, H.A.H., Reinders, I.: Experiences with using persuasive technology in a diet trial for older adults. In: In The 14th PErvasive Technologies Related to Assistive Environments Conf. (2021). https://doi.org/10.1145/3453892.3458686

[17] van der Lubbe, L.M., Klein, M.C.: Designing a system with persuasive communication to improve diet compliance for elderly users. In: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. pp. 234–241 (2019). https://doi.org/10.1145/3329189.3329217

[18] Ornish, D., Brown, S.E., Billings, J., Scherwitz, L., Armstrong, W.T., Ports, T.A., McLanahan, S.M., Kirkeeide, R.L., Gould, K., Brand, R.: Can lifestyle changes reverse coronary heart disease?: The lifestyle heart trial. The Lancet **336**(8708), 129–133 (1990). https://doi.org/10.1145/2792838.2799665

[19] Reinders, I., Wijnhoven, H.A.H., Jyväkorpi, S.K., Suominen, M.H., Niskanen, R., Bosmans, J.E., Brouwer, I.A., Fluitman, K.S., Klein, M.C.A., Kuijper, L.D., van der Lubbe, L.M., Olthof, M.R., Pitkälä, K.H., Vijlbrief, R., Visser, M.: Effectiveness and cost-effectiveness of personalised dietary advice aiming at increasing protein intake on physical functioning in community-dwelling older adults with lower habitual protein intake. BMJ Open **10**(11) (2020). https://doi.org/10.1136/bmjopen-2020-040637

[20] Zheng, Z., Kohavi, R., Mason, L.: Real world performance of association rule algorithms. In: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 401–406 (2001). https://doi.org/10.1145/502512.502572

# Part V

# Empower young adults to increase their mental well-being

# Introduction

## Motivation

Young adulthood is a critical developmental period, with different possible sources of stress. Experiences during these years have long-lasting implications on many aspects in later adult life [2]. For example, mental health issues developed during this period are often carried into adulthood without being treated. Self-compassion is a conceptualization of a healthy attitude and relationship to oneself, and can simply be explained as compassion turned inward [8]. It is shown that self-compassion is positively correlated with different measures of psychological health [10], and negatively correlated with psychopathology [6, 7]. Self-compassion applies to many different aspects of life, and many aspects of the mental well-being of young adults can benefit from it. Since the overall mental health of young adults is often challenged in different ways, learning about how to apply self-compassion can empower young adults to increase their mental well-being.

Self-compassion training programs often include the in-person guidance of one or more trainers that are present during the training. A well-known program to increase self-compassion is the 8-week Mindful Self-Compassion (MSC) Program [9]. However, young adults experience different barriers when it comes to seeking help for such problems, among which are a lack of knowledge about existing help, a lack of trust, and a preference for self-help [3]. In-person programs might therefore not suit how they would like to receive help. Regardless of this specific target group, in-person training programs have some practical disadvantages. First of all, the training programs often expect participants to be at a physical location. Moreover, one trainer can only facilitate one session at the same time, with a limited number of participants included. Finally, participants are required to join a group of strangers, and some participants might like to practice with this in the privacy and safety of their private environment and/or alone.

To overcome the burdens to join a program, and increase the scalability, online interventions can be a solution. A typical approach can be to create a similar program in which a trainer provides trainees with materials to work on, sometimes even including online group sessions. However, this still limits the scalability and can include a group component. Thus, it is interesting to develop an approach in which the training is completely self-guided. Creating such a self-guided intervention brings various challenges. The content of the training program needs to be adapted to the self-guided nature. Moreover, self-guided does not mean that the training program cannot be interactive. Interactive elements can be enhanced by AI techniques to make them more adaptive. Besides the content, also the form in which the content is delivered is important. One drawback of a self-guided program can be that participants might have difficulties with their motivation, as an external motivator, such as the supervision of a human trainer, is lacking. Gamification is a way to create another external motivation but at the same time it allows you to deliver content more interestingly [1].

## Approach

This part focuses on the following subquestion:

> *How can gamification be used to develop a self-guided intervention to enhance self-compassion and increase the well-being of young adults?*

In this part, I explore how to build such a self-guided intervention. To make it easily accessible and maintainable, the training is built in the form of a responsive website. For the content of the training, the aforementioned MSC Program is used as a guideline. Since this is an in-person program it is important to carefully think about how the practices from this program can be used in a self-guided context. With gamification elements, users are motivated to interact with the intervention and retain the program [1, 5, 4]. This can be achieved by rewarding users for completing certain tasks. Moreover, there is a wide variety of game elements that can be used to make the experience of the intervention better and more engaging [4]. An important aspect that I took into account in the design of the intervention is to ensure that different game elements are aligned in some way. With this, I aim to create a sense of unity in the intervention and make clear how different components relate to each other. All game elements are therefore related to one metaphor for increasing self-compassion, namely travelling through Europe. Another important aspect to focus on for a self-guided intervention in this domain is that it is safe for users' mental health. Since no professional is reading along, it is important to make sure that users are not likely to get mentally distressed by the training. A positive approach in both the content and the gamification is a good approach to dealing with this potential risk. The initial design of the intervention can be found in Chapter 8.

After a pilot evaluation with 7 participants, described in Chapter 9, the initial version is improved and evaluated in an experimental evaluation study. During this 6 weeks study, 294 participants were offered the training program and asked to fill in different questionnaires. The results on the user experience during this study is described in Chapter 11. It is very important to thoroughly evaluate how participants interacted with and evaluated (the interaction with) the website. Because the intervention is self-guided, participants must have a positive user experience to be able to successfully absorb the training content.

A challenge for a self-guided intervention is to create a certain level of personal adaptation in the exercises. To facilitate this, different AI components are used. First of all, sentiment analysis can be used in specific exercises to measure the sentiment in the answer of a user and adapt the reaction of the system following that sentiment. For this, I have used a sentiment analysis library. In addition to this, I have worked on a topic modelling algorithm that can assign topics to texts entered on the website by users. Different exercises use situation descriptions that are randomly chosen every time you want to practise. With this algorithm, I can tailor these situations to the personal experience of the user. Once this algorithm is added to the website, participants are asked whether they want a situation from a topic they have previously discussed or not. Details on the algorithm and its performance are described in Chapter 10.

## Position according to Chapter 2

In this section, I position the work of this part with the use of the different terms defined in Chapter 2, see Figure 2.1 and Figure 2.2. The terms are written in italics.

For this project, the target group is an *age group*. The young adults targeted by this training are vulnerable to a *health risk*, due to various circumstances in their life. The empowerment method that is used in this application is *training skills*; by using the application users learn how to practice self-compassion. This also includes knowledge transfer, but the main goal of that is being able to train practising it.

The application is available on the *web*, to make it platform-independent and easy to access. Different game elements are used to motivate participants but also to deliver the knowledge more engagingly. *Objectives*, in the form of intermediate point (kilometre) goals and streaks, are used to motivate participants to keep on using the intervention. Moreover, *rewards* are given when a participant completes an exercise: points in the form of kilometres are awarded when a participant completes a part of the intervention. However, there is a maximum number of kilometres a day to discourage excessive use. Moreover, checkmarks are placed at each intervention part that has been completed on a day. Players can *get insight into their own data* as they can see their previous entries for each intervention part. The theory is delivered in the form of a *story*. In this story, users have some possibilities for *customization*, such as choosing the next destination in the story.

A first version of the website is evaluated in a *pilot study* with *7 participants*, whom both filled in questionnaires but also got interviewed (one participant did not get interviewed). Moreover, an expert working on a similar intervention was interviewed. Although overall the results were *positive*, different points for improvements were reported. Based on these results the training website is improved and evaluated in a *6-weeks experimental study*. In total *583 participants* were included and divided over the training condition (*294 participants*) and the control condition (*289 participants*). The results on the user experience show a *high drop-out*, which largely consists of participants that dropped out before any interaction or only some interaction in the first week. Overall, *completers are actively engaged with the training and evaluate it positively*, although comments are made about improving mainly the clarity of the website and training program and offering different forms of content.

## This part contains:

# Bibliography


[1] AlMarshedi, A., Wanick, V., Wills, G.B., Ranchhod, A.: Gamification and Behaviour, pp. 19–29. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-45557-0_2

[2] Council, N.R., et al.: Investing in the health and well-being of young adults (2015)

[3] Gulliver, A., Griffiths, K.M., Christensen, H.: Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC psychiatry **10**(1), 1–9 (2010)

[4] Kelders, S.M., Sommers-Spijkerman, M., Goldberg, J.: Investigating the direct impact of a gamified versus nongamified well-being intervention: An exploratory experiment. J Med Internet Res **20**(7), e247 (Jul 2018). https://doi.org/10.2196/jmir.9923

[5] Litvin, S., Saunders, R., Maier, M.A., Lüttke, S.: Gamification as an approach to improve resilience and reduce attrition in mobile mental health interventions: A randomized controlled trial. PloS one **15**(9), e0237220 (2020)

[6] MacBeth, A., Gumley, A.: Exploring compassion: A meta-analysis of the association between self-compassion and psychopathology. Clinical Psychology Review **32**(6), 545–552 (2012). https://doi.org/https://doi.org/10.1016/j.cpr.2012.06.003

[7] Marsh, I.C., Chan, S.W., MacBeth, A.: Self-compassion and psychological distress in adolescents—a meta-analysis. Mindfulness **9**(4), 1011–1027 (2018)

[8] NEFF, K.: Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. Self and Identity **2**(2), 85–101 (2003). https://doi.org/10.1080/15298860309032

[9] Neff, K.D., Germer, C.K.: A pilot study and randomized controlled trial of the mindful self-compassion program. Journal of Clinical Psychology **69**(1), 28–44 (2013). https://doi.org/https://doi.org/10.1002/jclp.21923

[10] Zessin, U., Dickhäuser, O., Garbade, S.: The relationship between self-compassion and well-being: A meta-analysis. Applied Psychology: Health and Well-Being **7**(3), 340–364 (2015). https://doi.org/https://doi.org/10.1111/aphw.12051

# Chapter 8

# Designing a gamified self-compassion training

This chapter is published as: **van der Lubbe, L.M.**, Gerritsen, C., Klein, M.C.A., Hindriks, K.V., & Rodgers, R.F. (2021). Designing a gamified self-compassion training. In A. Veloso, O. Mealha, & L. Costa (Eds.), 22nd International Conference on Intelligent Games and Simulation, GAME-ON 2021 (pp. 79-85). Eurosis.

**This chapter contains:**



8

## Abstract


Social media platforms have become increasingly visual. This has been associated with body image issues, especially for young adults who are frequent users of such platforms. Increasing ones self-compassion can mitigate these potential negative effects. In this chapter we present a training for self-compassion that young adults can follow online. The training is self-guided and individual, and users can use it at any preferred moment. Moreover, we use gamification in the training to motivate users and to deliver theory in

an engaging way. In this chapter we describe the theoretical background and the choices that have been made in the design of the training.

## 8.1 Introduction

Young adults use more so-called highly visual social media (like Snapchat[1] and Instagram[2]), compared to Facebook[3]. The use of these highly visual social media is related to body image concerns and internalizing symptoms [16, 26]. Although this research is very recent, new types of social media that are also highly visual have emerged since then, for example TikTok[4], and may have even greater detrimental effects on body image. Body image is an important concern among young adults, and is associated with a range of negative outcomes such as depression and eating disorders [29]. The negative effects on body image and mood of engagement with high-visual social media can largely me accounted for by social comparison tendencies, including appearance-based comparisons as well as comparisons on other dimensions [15, 26]. One of the dimensions that has been suggested to be protective against social comparison is self-compassion.

Self-compassion is an attitude of observant and non-judgmental self-kindness and has been associated with a range of positive outcomes and psychological well-being [17]. In the context of body image, self-compassion may not only decrease social comparison but also encourage an attitude of self-acceptance, and the recognition that social standards of attractiveness are unrealistic for most individuals through healthy means [1]. Given this and preliminary work among adolescents in this area [23], it is likely that an intervention targeting self-compassion in the context of social media use would be helpful to protect young adults against body image concerns.

In this chapter the design of an individual online training for self-compassion is described. First, background literature is presented in more detail. Next, the design of the training is discussed. To conclude, we briefly discuss the future work of this project.

## 8.2 Background

### 8.2.1 Self-ccompassion

Self-compassion consists of three components [17]. First is self-kindness, meaning being kind and understanding towards oneself instead of being self-critical. Next, common humanity refers to perceiving your experiences as part of a larger human experience instead of being isolated. Finally, mindfulness involves a balanced view on negative emotions, instead of over-identifying with them.

As mentioned above, self-compassion has been shown to be related to lower levels of social comparison [21], one of the mechanisms accounting for the detrimental effects

[1] https://www.snapchat.com/
[2] https://www.instagram.com/
[3] https://www.facebook.com/
[4] https://www.tiktok.com/

of media and social media on body image and mood. In addition, self-compassion has been shown to weaken the relationships between perceived higher weight and appearance comparison [24].

Self-compassion has been related to many positive psychological outcomes [17, 21]. For example, it reduces how much people suffer from a negative event. Self-compassion may contribute to effective self-regulation and it can cause people to have a more positive way of handling shortcomings. Moreover, a higher level of self-compassion helps to handle stress in a more ‘problem-focused’ way. Self-compassion is also a negative predictor for social comparison.

Self-compassion has been correlated with lower anxiety and depression [19, 11], and influences may render individuals less prone to experiencing shame [11]. It has been correlated with higher well-being, life satisfaction, happiness, optimism, emotional intelligence, and use of adaptive coping strategies [21, 19, 11], and it has also been related to greater social connection [19]. Furthermore, it has shown a relationship with higher agreeableness, a lower desire to please and a higher ability to say no [20].

Several interventions have been shown to be successful in increasing self-compassion. The 8-week Mindful Self-Compassion (MSC) Program is a program aiming to increase participant’ self-compassion. In a randomized controlled trial, participating with the MSC Program was found to increase self-compassion more compared to the control group, which was maintained at 6- and 12-months followups [18]. In other programs, such as the Mindfulness-Based Stress Reduction (MBSR) program [12] and Mindfulness-Based Cognitive Therapy (MBCT) [28], self-compassion is not a principal target and changes are a byproduct of the focus on one of its components; here mindfulness. An adapted version of the MSC Program called ‘making friends with yourself’ was created to target adolescents. This program lasts six weeks, each week having its own theme for the scheduled session. During a pilot study this program was found to be effective in increasing self-compassion [4]. Self-compassion training can also be facilitated in other ways. For example, in a study with young adults, support was found for the feasibility for a video-conference intervention [6]. This research specifically targeted young adult cancer survivors. Another app aims to help cancer patients from different ages with learning about self-compassion [2]. Although overall it is shown that self-compassion can be increased with the MSC Program, it has not been used for body image concerns related to social media usage.



### 8.2.2 Body image

Body image concerns, including body dissatisfaction, refer to negative self-evaluations regarding appearance associated with thoughts, feelings, and behaviours. Body image concerns are a critical risk factor for disordered eating [29]. Sociocultural theories of body image highlight how the promotion of unattainable appearance ideals leads to their adoption by individuals as a personal standard. In turn, this is associated with body dissatisfaction as appearance ideals are unrealistic and unattainable for most individuals through healthy means [30].

As briefly described above, highly-visual social media has recently emerged as an important influence on body image [26]. Social media used by young adults is nowadays highly visual and presents highly curated self-presentations that aim to display idealized bodies and lifestyles. Programs like Photoshop or filters are used to enhance pictures before posting. Moreover, in addition to following friends and acquaintances, following high-profile individuals with whom one does not have a personal relationship (such as influencers and celebrities) is an important part of using social media. Using highly-visual social media, engaging in appearance comparisons, and having high degrees of emotional investment in one's own online self-presentation have been shown to be associated with negative outcomes for young people [15, 31]. In addition to the detrimental effects on body image, engaging with high-visual social media has also been associated with negative affect and depression among youth [16].

Self-compassion has started to emerge as a potential protective dimension against body image concerns. A systematic review showed that self-compassion was associated with lower eating pathology and improved body image across studies [5]. In addition, some evidence has emerged for self-compassion as a moderator between risk factors and body image and eating concerns. Thus, self-compassion was found to mediate the relationship between external shame and drive for thinness [8], and body image disturbance and psychological distress [22]. Moreover, self-compassion was found to moderate the association between the appearance comparison and lower body appreciation [10].

In this way, self-compassion has been associated with more positive body image and lower levels of body image concerns and related risk factors, and therefore may be a useful tool for young people in the context of social media use. The three components of self-compassion may each play a protective role. Self-kindness is inherently inconsistent with being critical towards the body, and may therefore help foster body appreciation and acceptance. Common humanity may be useful for remembering that online portrays are polished and unrealistic, and that most individuals feel that they do not live up to idealized social media self-presentations. Finally, mindfulness may help to reduce negative body-related thoughts and emotions by helping to put them in perspective [1]. In addition, self-compassion may reduce pressure to hold oneself to external standards, thereby reducing social comparison [3].

Building on this, preliminary support has been found for the usefulness of self-compassion based interventions for body image. Among adult women, a 3-week intervention with a self-compassion meditation podcast was successful in improving self-compassion and body image outcomes [1]. Furthermore, findings revealed that the intervention was also effective in reducing body dissatisfaction, body shame, contingent self-worth based on appearance. These gains were maintained at 3 months. In another study, mobile intervention called BodiMojo designed for late adolescents and emerging adults revealed promising results for self-compassion and body image and related aspects in a six-week randomized controlled trial [23]. These preliminary findings support further examination of the efficacy of self-compassion based mobile or online interventions.

### 8.2.3 Gamification and body image

Gamification can be defined as 'the intentional use of game elements for a gameful experience of non-game tasks and contexts' [27]. In the context of mental health applications, it has the potential to increase positive outcomes in different ways: by appealing to users, through its potential reach, by increasing engagement, and by its capacity to target multiple mechanisms of change [9]. The aforementioned mobile intervention for self-compassion (BodiMojo) did not include gamification, and to the best of our knowledge no interventions targeting self-compassion using gamification have been developed or evaluated. Gamified interventions have been developed for body image issues more broadly. For example, a virtual reality game, which can be used for eating disorders treatment [25]. Other examples include app-based interventions, using gamification, aiming to decrease body dissatisfaction and related risks [7, 13].

## 8.3 Designing an independent digital self-compassion training

The target group for our training was young adults, who are vulnerable to having body image related issues which are (partly) due to their social media usage. The aim of the intervention was to promote self-compassion among young adults, to make this group more resilient to body image threats in the context of social media use. Fortunately, learning more about self-compassion also enhances other factors related to well-being, as explained in the background section. To make the training accessible to a broad target group, it was made available on different platforms. A responsive website made it possible to use the training on different types of devices, independent of the operating system. Although an installable mobile app would have some benefits over using a website, such as sending push notifications, it was more convenient to use a website for this project. Mobile apps require platform specific codes, while websites are suitable for all different platforms without these different coding languages.



### 8.3.1 Design of the intervention

Many online interventions include social elements to enhance the user experience. However, for this intervention it was important that users felt safe to share their thoughts freely. Therefore, an individual user experience with no interactive components was chosen. Moreover, the intervention was independent, meaning no professional or experts were available to interact with users. It was important to make users aware of this, upfront but also during the intervention. Information regarding (online) resources for professional help-seeking was available on all pages of the website, in case of need.

The website included of five components: a profile, exercises, a journal, gamification, and motivational messages. The profile was mainly administrative for users to see some of their usage statics, as well as a way to change some personal settings for the website. The exercises and the journal together formed the actual training.

### 8.3.2 Exercises

The training included three exercises, based on exercises from the MSC Program [18]. In the MSC Program, the exercises are guided by the trainer. In our intervention, the exercises were guided by the system. They were designed in a lay-out of text messages, to give it the feel of a text conversation, see Figure 8.1. To make users write extensive

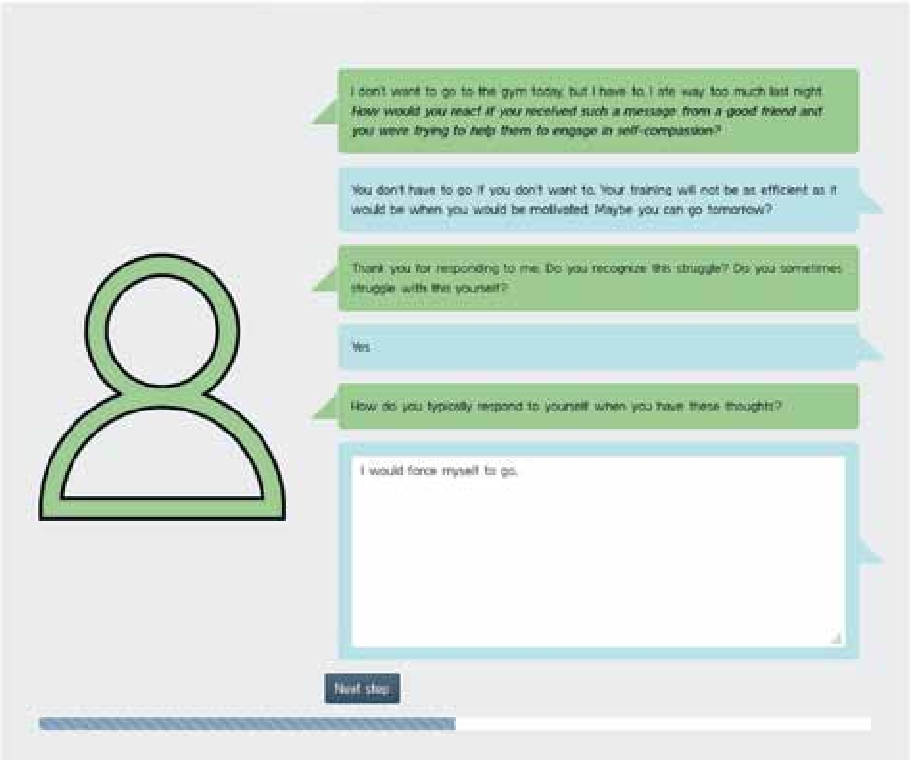



**Figure 8.1:** Screenshot of Exercise 1

texts without any personal guidance of a trainer, we chose to change the exercises to focus on individual practice. Moreover, this approach reduced the time that users spent on the exercise, which was assumed to make it less burdensome. Moreover, text message styled conversations are recognisable for the envisioned target group. For all text written by users in the exercises, a minimum of 10 characters was specified to enforce at least some elaboration by the users. With user testing later in this project we will measure the average number of characters in an answer and further specify this minimum.

#### 8.3.2.1 Exercise 1: How would I treat a friend?

This exercise from the MSC program focuses on the fact that self-compassion can be likened to being a good friend to oneself. First, users had to think about how they would support a friend in difficult times. Next, they had to think about how they support themselves in difficult times, see Figure 8.1. Finally, users reflected on the differences in how they treat their friend and how they treat themselves. The exercise in our intervention took the same steps. For the exercise a randomly chosen situation was described in which a friend shares a concern with you. In the MSC Program the trainer can anticipate on the answers of the participants, we tried to recreate this with the use of sentiment analysis (using the Vader Library for Python[5]). The final prompt was tailored to whether, based on the sentiment analysis of the first two steps, the user is more positive or negative towards themself compared to the imagined friend.

#### 8.3.2.2 Exercise 2: Compassionate message

This exercise was based on the exercise 'compassionate letter to myself' from the MSC Program, in which users write a letter to themselves in a compassionate voice. In the original MSC program, different approaches to this exercise are offered: (1) write from the compassionate self to the struggling self ('me to me'), (2) write from the compassionate other to oneself ('you to me'), or (3) write from the compassionate self to another person ('me to you'). In our training, one of these approaches was chosen randomly every time the exercise started. Moreover, sometimes the exercise also proposed a problem that the user should address in the letter. Instead of having to write an extensive letter, users were asked to write a message which did not need to be very lengthy, but has to meet the character minimum.

#### 8.3.2.3 Exercise 3: Practising self-compassion

This exercise was based on the exercise 'motivate ourselves with compassion' from the MSC Program, and is about motivating oneself with self-compassion instead of self-criticism. The original steps of the exercise are:

1. Thinking about a behaviour that you want to change;
2. Finding your self-critical voice;
3. Compassion for feeling criticised;
4. Turning toward your inner critic;
5. Finding your compassionate voice.

The exercise in our training was shortened to include only the first and last step. This choice was made to reduce the time that was spent on the exercise, but, more importantly, because this focuses more on the positive aspects of the exercise instead of the negative ones. Without personal guidance, this is a safer choice, as this avoids users to dwell on their negative thoughts.

### 8.3.3 Journal

The training included a private journal for the user. In this journal, the user could make notes about three different aspects: moods, gratitude and using self-compassion. Mood and gratitude could only be recorded once a day in the evening.

[5] https://pypi.org/project/vaderSentiment/

#### 8.3.3.1 Mood journal

In the mood journal, the users monitored their moods. This stimulates the skill of emotional awareness, which is part of the mindfulness component of self-compassion. To monitor their mood, users could choose from five different basic emotions: afraid, angry, ashamed, happy, and sad. Next, the user needed to choose an intensity for that mood: low, medium or high. Examples of words describing that intensity of the mood were shown to help users find the right level of intensity. Graphically, the basic moods were represented by smileys with different colours and expressions. For the different intensities, the same smileys were used, but the intensity of the colour was reduced, see Figure 8.2.

In addition to registering their moods, a user could also view their previous moods. Two different overviews were provided to the user. The first was a list of all the smileys that the user had chosen in chronological order. This gave the user a sense of how their mood was changing or stable over time. The other overview showed per mood how often the user registered the mood. With this overview the user could see which was the most/less frequent moods that (s)he noted in the journal.

Importantly, in this journal no value was given to specific moods, to avoid suggesting some moods were better compared to other moods. For example, the moods were ordered alphabetically when choosing a mood to register. Moreover, users were asked to focus more on the overall mood of the day, instead of being able to log all (small) mood changes.

#### 8.3.3.2 Gratitude journal

One of the informal practices described in the MSC Program is about counting blessings. Participants were asked to write down small things that they felt grateful for. In the gratitude journal in this intervention, users could write down something that they currently felt, or recently felt, grateful for. Moreover, users could give a tag to their note. This tag was stored


**Figure 8.2:** Screenshot of the mood journal

without spaces and a hashtag at the beginning, to represent hashtags used on social media. Hashtags are familiar to the proposed target group. Previously used (hash)tags appeared to the user as suggestions when writing a new journal note, but as many new tags as the user wanted could be added. Gratitude is related to different benefits, among which is emotional well-being [32]. As with all components of our training, gratitude also had a positive focus.

As an overview of this journal was created through a tag cloud, see Figure 8.3. This was a list of all the tags, in which the size of the tag was determined by the frequency of use. The most often used tag(s) appeared in the largest and most prominent font. By clicking on a tag, the user could read all the notes that they had written for that tag.


**Figure 8.3:** Example of the gratitude journal's wordcloud

#### 8.3.3.3 Self-compassion journal

To give users a journal to keep track of their experiences with self-compassion, the self-compassion journal was included. The original MSC Program includes a session in which participants respond to the question 'how is MSC going for me?'. Although the self-compassion journal was less direct in this, it was a way for users to see their progress, but also learn to see the (potential) role of self-compassion in their life.

In this journal users could write about situations in which they applied, or would have liked to apply, self-compassion. Writing in this journal was divided in three steps. First the user chose which of the two mentioned types of situations they wanted to write about. Next, the user gave a description of the situation, together with a tag, similar to the gratitude journal. In the final step, the user was asked to explain more about the application of self-compassion in that situation. Examples were given to provide inspiration. When writing about a situation in which self-compassion was applied, the user was also asked to rate the effect on a 5-star scale.

To browse their self-compassion journal, users were offered different overviews. Comparable to the gratitude journal, a tag cloud was created for the tags used for the notes. In addition, two separate tag clouds were created for the two different types of notes. Users could also see the number of the different star ratings they gave to their notes in a bar chart, and a line chart is added to show them how the ratings differ over time.

### 8.3.4 Gamification

To engage users more in using the system, gamification was also included. The main gamification element was narrative. Theory about self-compassion was provided to the user with the help of a story. After signing up, a tutorial was provided that introduced the main story of the intervention: you were going on a trip through Europe, and on the first part of your trip you met an old friend. This old friend told you about his/her previous trip, in which (s)he was introduced to self-compassion. The story introduced all components of the training, and restricted users to try everything once. After that, all components of the website were unlocked.

During the tutorial the user could choose their departure city and first destination from a list of cities. Destinations could be chosen from a list of the five 'closest' cities that has not yet been visited. During the training, the player earned 25 kilometres for each exercise and journal entry. Once all components were completed at least once on a day, the player received 50 bonus kilometres. The maximum kilometres travelled per day was 300, to encourage users to use the website moderately over time, as fewer benefits are likely to result from intense use over a short time period. Once a user reached their destination, a notification was sent with another part of the story of your European trip. In each of these story parts, a situation was described in which self-compassion was applied. Moreover, the role of self-compassion was also explained explicitly, to generalize from the specific situation.

Reaching a new destination represented an achievement. The user could find their trip on the 'passport page' where the tutorial and destination texts could be found and re-read. Customization was used to engage users more in the story. Users could choose how their accompanying friend was called and users chose their own departure city and destination.

### 8.3.5 Motivational messages

In addition to the above mentioned way of delivering more theory about self-compassion to the user, motivational messages were also added to the homepage. These messages changed from time to time and included motivational messages about the topic of self-compassion, social media or the intervention.

## 8.4 Future work

In this chapter we presented the background and design of an independent, online, self-compassion training for young adults. The aim of the training was to increase resiliency against the documented negative effects of social media on body image and mood among young adults. To design this training, we examined existing programs for self-compassion and tried to include active components in the type of training we envisioned. The designed training consisted of three exercises and a journal with three different components. We ensured that the design of the exercise or journal fit the independent self-guided character of the training. Therefore the intervention largely focused on positive thoughts and elements throughout the exercises and journals, rather than inviting users to dwell on negative experiences. After the design phase, we performed a pilot study with participants from the target group [14]. In this pilot, interviews and questionnaires about their experiences gave us directions to further improve the system before we evaluate its effectiveness and user-experience in an experimental design. This chapter provides the theoretical foundation of the training and shows how traditionally offline training can be translated into a digital version.

## Bibliography

[1] Albertson, E.R., Neff, K.D., Dill-Shackleford, K.E.: Self-compassion and body dissatisfaction in women: A randomized controlled trial of a brief meditation intervention. Mindfulness **6**(3), 444–454 (2015). https://doi.org/10.1007/s12671-014-0277-3

[2] Austin, J., Drossaert, S.C., van Dijk, J., Sanderman, R., Schroevers, M.J., Bohlmeijer, E.T.: Self-compassion by tech: A demonstration of a supportive self-compassion app for people with newly diagnosed cancer. In: Supporting Health by Technology 2020 (2020)

[3] Berry, K.A., Kowalski, K.C., Ferguson, L.J., McHugh, T.L.F.: An empirical phenomenology of young adult women exercisers' body self-compassion. Qualitative research in sport and exercise **2**(3), 293–312 (2010). https://doi.org/10.1080/19398441.2010.517035

[4] Bluth, K., Gaylord, S.A., Campo, R.A., Mullarkey, M.C., Hobbs, L.: Making friends with yourself: A mixed methods pilot study of a mindful self-compassion program for adolescents. Mindfulness **7**(2), 479–492 (2016). https://doi.org/10.1007/s12671-015-0476-6

[5] Braun, T.D., Park, C.L., Gorin, A.: Self-compassion, body image, and disordered eating: A review of the literature. Body Image **17**, 117–131 (2016). https://doi.org/10.1016/j.bodyim.2016.03.003

[6] Campo, R.A., Bluth, K., Santacroce, S.J., Knapik, S., Tan, J., Gold, S., Philips, K., Gaylord, S., Asher, G.N.: A mindful self-compassion videoconference intervention for nationally recruited posttreatment young adult cancer survivors: feasibility, acceptability, and psychosocial outcomes. Supportive Care in Cancer **25**(6), 1759–1768 (2017). https://doi.org/10.1007/s00520-017-3586-y

[7] Cerea, S., Ghisi, M., Bottesi, G., Manoli, T., Carraro, E., Doron, G.: Cognitive behavioral training using a mobile application reduces body image-related symptoms in high-risk female university students: A randomized controlled study. Behavior Therapy **52**(1), 170–182 (2021). https://doi.org/10.1016/j.beth.2020.04.002

[8] Ferreira, C., Pinto-Gouveia, J., Duarte, C.: Self-compassion in the face of shame and body image dissatisfaction: Implications for eating disorders. Eating Behaviors **14**(2), 207–210 (2013). https://doi.org/10.1016/j.eatbeh.2013.01.005

[9] Fleming, T.M., Bavin, L., Stasiak, K., Hermansson-Webb, E., Merry, S.N., Cheek, C., Lucassen, M., Lau, H.M., Pollmuller, B., Hetrick, S.: Serious games and gamification for men-

8

tal health: current status and promising directions. Frontiers in psychiatry **7**, 215 (2017). https://doi.org/10.3389/fpsyt.2016.00215

[10] Homan, K.J., Tylka, T.L.: Self-compassion moderates body comparison and appearance self-worth's inverse relationships with body appreciation. Body Image **15**, 1–7 (2015). https://doi.org/10.1016/j.bodyim.2015.04.007

[11] Johnson, E.A., O'Brien, K.A.: Self-compassion soothes the savage ego-threat system: Effects on negative affect, shame, rumination, and depressive symptoms. Journal of Social and Clinical Psychology **32**(9), 939–963 (2013). https://doi.org/10.1521/jscp.2013.32.9.939

[12] Kabat-Zinn, J., Hanh, T.N.: Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness. Delta (2009)

[13] Kollei, I., Lukas, C.A., Loeber, S., Berking, M.: An app-based blended intervention to reduce body dissatisfaction: A randomized controlled pilot study. Journal of consulting and clinical psychology **85**(11), 1104 (2017). https://doi.org/10.1037/ccp0000246

[14] van der Lubbe, L., Gerritsen, C., Klein, M., Hindriks, K., Rodgers, R.: A pilot study of a gamified self-compassion training. Proceedings of the 22nd European GAME-ON conference (2021), to appear

[15] Lup, K., Trub, L., Rosenthal, L.: Instagram #instasad?: Exploring associations among instagram use, depressive symptoms, negative social comparison, and strangers followed. Cyberpsychology, Behavior, and Social Networking **18**(5), 247–252 (2015). https://doi.org/10.1089/cyber.2014.0560, pMID: 25965859

[16] Marengo, D., Longobardi, C., Fabris, M., Settanni, M.: Highly-visual social media and internalizing symptoms in adolescence: The mediating role of body image concerns. Computers in Human Behavior **82**, 63–69 (2018). https://doi.org/10.1016/j.chb.2018.01.003

[17] Neff, K.: Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. Self and Identity **2**(2), 85–101 (2003). https://doi.org/10.1080/15298860309032

[18] Neff, K.D., Germer, C.K.: A pilot study and randomized controlled trial of the mindful self-compassion program. Journal of clinical psychology **69**(1), 28–44 (2013). https://doi.org/10.1037/0022-3514.92.5.887

[19] Neff, K.D., Kirkpatrick, K.L., Rude, S.S.: Self-compassion and adaptive psychological functioning. Journal of Research in Personality **41**(1), 139–154 (2007). https://doi.org/10.1016/j.jrp.2006.03.004

[20] Neff, K.D., Rude, S.S., Kirkpatrick, K.L.: An examination of self-compassion in relation to positive psychological functioning and personality traits. Journal of Research in Personality **41**(4), 908–916 (2007). https://doi.org/10.1016/j.jrp.2006.08.002

[21] Neff, K.D., Vonk, R.: Self-compassion versus global self-esteem: Two different ways of relating to oneself. Journal of Personality **77**(1), 23–50 (2009). https://doi.org/10.1111/j.1467-6494.2008.00537.x

[22] Przezdziecki, A., Sherman, K.A., Baillie, A., Taylor, A., Foley, E., Stalgis-Bilinski, K.: My changed body: breast cancer, body image, distress and self-compassion. Psycho-oncology **22**(8), 1872–1879 (2013). https://doi.org/10.1002/pon.3230

[23] Rodgers, R.F., Donovan, E., Cousineau, T., Yates, K., McGowan, K., Cook, E., Lowy, A.S., Franko, D.L.: Bodimojo: Efficacy of a mobile-based intervention in improving body image and

self-compassion among adolescents. Journal of youth and adolescence **47**(7), 1363–1372 (2018). https://doi.org/10.1007/s10964-017-0804-3

[24] Rodgers, R.F., Franko, D.L., Donovan, E., Cousineau, T., Yates, K., McGowan, K., Cook, E., Lowy, A.S.: Body image in emerging adults: The protective role of self-compassion. Body Image **22**, 148–155 (2017). https://doi.org/10.1016/j.bodyim.2017.07.003

[25] Rodrigues, F., Wilbert, P., da Silva, J.C.T., de Oliveira, J.C.: A virtual reality environment using concepts of serious games and gamification for the treatment of eating disorders. In: Anais do XIX Simpósio Brasileiro de Computação Aplicada à Saúde. pp. 141–151. SBC (2019). https://doi.org/10.5753/sbcas.2019.6249.

[26] Saiphoo, A.N., Vahedi, Z.: A meta-analytic review of the relationship between social media use and body image disturbance. Computers in Human Behavior **101**, 259–275 (2019). https://doi.org/10.1016/j.chb.2019.07.028

[27] Seaborn, K., Fels, D.I.: Gamification in theory and action: A survey. International Journal of Human-Computer Studies **74**, 14–31 (2015). https://doi.org/10.1016/j.ijhcs.2014.09.006

[28] Segal, Z.V., Williams, M., Teasdale, J.: Mindfulness-based cognitive therapy for depression. Guilford Publications (2018)

[29] Stice, E., Marti, C.N., Durant, S.: Risk factors for onset of eating disorders: Evidence of multiple risk pathways from an 8-year prospective study. Behaviour Research and Therapy **49**(10), 622–627 (2011). https://doi.org/10.1016/j.brat.2011.06.009

[30] Thompson, J.K., Heinberg, L.J., Altabe, M., Tantleff-Dunn, S.: Exacting beauty: Theory, assessment, and treatment of body image disturbance. American Psychological Association (1999). https://doi.org/10.1037/10312-000

[31] Veldhuis, J., Alleva, J.M., Bij de Vaate, A.J., Keijer, M., Konijn, E.A.: Me, my selfie, and i: The relations between selfie behaviors, body image, self-objectification, and self-esteem in young women. Psychology of Popular Media **9**(1), 3 (2020). https://doi.org/10.1037/ppm0000206

[32] Wood, A.M., Froh, J.J., Geraghty, A.W.: Gratitude and well-being: A review and theoretical integration. Clinical Psychology Review **30**(7), 890–905 (2010). https://doi.org/10.1016/j.cpr.2010.03.005, positive Clinical Psychology

# Chapter 9

# A pilot study of a gamified self-compassion training

This chapter is published as: **van der Lubbe, L.M.**, Gerritsen, C., Klein, M.C.A., Hindriks, K.V., & Rodgers, R.F. (2021). A pilot study of a gamified self-compassion training. In A. Veloso, O. Mealha, & L. Costa (Eds.), 22nd International Conference on Intelligent Games and Simulation, GAME-ON 2021 (pp. 86-93). Eurosis.

**This chapter contains:**



9

## Abstract

The rates of mental health concerns among young adults are high. Self-compassion has been shown to hold benefits for mental health. This chapter presents a pilot study of an online gamified training for young adults designed to increase self-compassion. Specifically, user experience and engagement were explored. Participants from the target group were invited to use the training for 2 or 6 weeks. User experience and engagement were assessed both quantitatively and qualitatively through questionnaires and exit interviews. Findings revealed that participants liked the practical approach to self-compassion of the training. Moreover, participants liked and were motivated by the gamification elements. In

addition, participants offered several points for improvement, including greater variety in the exercises, more cheerful colours for the website, and the use of more story elements.

## 9.1 Introduction

The rates of mental health concerns among young adults are high. One potential contributor to these concerns includes the use of highly visual social media, such as Instagram[1] or TikTok[2]. Such highly visual social media have been related in particular to body image issues [6]. This is concerning as in addition to being highly distressing, body image disturbance is a known risk-factor for eating disorders, depression, and low self-esteem [16].

Self-compassion has been emerging as an important psychological dimension associated with positive outcomes [9], including positive body image. To date, most interventions targeting self-compassion are designed to be conducted in person, either individually or in a group [2]. This may constitute barriers to engaging in such interventions. The aim here was therefore to create a self-guided online, gamified, self-compassion training for young adults. The training included exercises and journals aiming to increase the users' skill in self-compassion specifically as applied to the effects of social media on their body image with the goal of promoting positive body image. Gamification was employed to increase user motivation and engagement through a narrative approach and rewards.

A pilot study was conducted to examine user experience and engagement. The findings from this pilot study will inform the design of the full training that will be tested in an experimental trial. In the next sections, the background literature and the training are introduced. Then, the study methodology and findings are presented and discussed with implications for future research.

## 9.2 Background

Self-compassion consists of three components [9]:

1. Self-kindness: being kind and understanding towards oneself instead of being self-critical;

2. Common humanity: perceiving your experiences as part of a larger human experience instead of being isolated;

3. Mindfulness: having a balanced view on negative emotions, instead of over-identifying with them.

Self-compassion has been associated with positive outcomes. For example, it has been shown to be associated with lower anxiety and depression, and greater well-being, life satisfaction, happiness, optimism, emotional intelligence, and positive coping strategies [11,

[1]https://www.instagram.com/
[2]https://www.tiktok.com/

4]. Several interventions have been shown to be successful in increasing self-compassion, for example the 8-week Mindful Self-Compassion (MSC) Program [10].

Gamification can be defined as 'the intentional use of game elements for a gameful experience of non-game tasks and contexts' [15]. It is often used to enhance users' motivation. In the context of mental health applications, adding gamification may increase the intervention's efficacy [3].

Serious games can also be used in the context of mental health. A serious game can be defined as: "A game in which education (in its various forms) is the primary goal, rather than entertainment" [8]. A review study showed that serious games and gamification were useful in interventions designed to target mental health concerns [17]. However, to date, no interventions using gamification for self-compassion exist.

## 9.3 Method

### 9.3.1 Design

Prior to evaluating the intervention in a full trial, a pilot study was conducted. The goal of the pilot study was to examine user experience and engagement with the training to inform the design of both the intervention and its evaluation in the larger trial. The pilot study aimed to include around 35 participants, for a duration of 2 or 6 weeks. Initially, the study was designed to last 6 weeks, however, recruitment challenges led to this being reduced to 2 weeks at the end of the inclusion phase. Thus, participants who started at the beginning of the inclusion phase used the system for 6 weeks, while participants included at the end used it for only 2 weeks. Both the intervention and all the assessments were conducted in Dutch.

In addition, we interviewed a Dutch researcher who was also working on a self-compassion app, but for a different target audience. We invited her to explore our training system, and participate in a structured interview afterwards. The goal of this interview was to obtain insights from this domain expert on the content and functionalities.



#### 9.3.1.1 Individual digital self-compassion training

A self-guided digital self-compassion training intervention for young adults was developed (Chapter 8, [5]). The training can be accessed via a website, where it is always available. The training is individual and self-guided, meaning there is no interaction component with other users or a professional.

Three exercises based on the MSC Program [10] were included in the training. While in the MSC Program the exercises are guided by a professional, the exercises on the website are guided by the system. Therefore, all exercises focused exclusively on positive elements, and none of the content was focused on negative emotions or experiences. The exercises are designed in the format of text message conversations, and feel like a conversation with the system, see Figure 9.1a. To ensure that users elaborate on their answers, a character minimum was set at 10 characters. In addition to exercises, the user could access a journal consisting of three components:

(a) Screenshot of an exercise


(b) Screenshot of the mood journal

(c) Screenshot of the journey progression and passport page

**Figure 9.1:** Screenshots from the training

- A mood journal to keep track of the users mood, and designed to target mood regulation, see Figure 9.1b;
- A gratitude journal to stimulate emotional well-being;
- A self-compassion journal to keep track of situations in which self-compassion (could have) helped the user.

In addition, a series of motivational messages about self-compassion that varied over time were shown on the first page of the website.

Gamification was used to engage the users more with the system, with the use of a narrative, goals and points as main game elements. The tutorial at the start of the training was written as a story about a journey the user embarks upon. Some aspects of the story could be customised including the name of the main character and the destinations. When engaging with the interventions, users earned kilometres for their journey. The progress on the journey was visible on every page in the form of a train on a progress bar. Figure 9.1c shows the progress bar on top. When reaching a new location (goal) the user received a short story about the destination which again included theoretical elements regarding self-compassion in a narrative form. This information remained accessible to users on their passport page, see Figure 9.1c. The reward kilometres were limited per day, to encourage regular but limited interactions with the system.

### 9.3.2 Measures

Before accessing the intervention, participants were asked to complete a series of self-report measures. Upon completing the intervention, (after 2 or 6 weeks) participants completed a user experience questionnaire and were invited to participate in a structured exit interview. Data were also collected from the website.

#### 9.3.2.1 Demographics

Demographic information was collected at the beginning of the study, including: age, gender, highest or current education level. Participants also reported whether they were currently receiving therapy for a mental health disorder. Finally, the term social media was introduced, and examples of platforms that were considered to constitute social media provided. Chat services like WhatsApp[3] or Telegram[4] were not considered to be social media because they are mainly used as personal communication service. Participants also estimated their daily social media use.

In addition, we included a questionnaire on photo investment [7], that captures the investment and effort that participants put in choosing photos to share online is measured. The score for this questionnaire is calculated by averaging all question scores; the higher the final score the higher the investment in photo sharing online. The short self-compassion scale was used to measure the level of self-compassion of the participant [13]. To calculate the self-compassion score, first the averages of each subscale were calculated, then the mean of the all subscales were taken. To measure self-objectification four items from the Body Surveillance Subscale of the Objectified Body Consciousness Scale for Preadolescent and Adolescent Youth, translated in Dutch [18] were used. A higher score indicates higher levels of self-objectification. Finally, body image flexibility was measured [14]. A higher average score indicates more body image flexibility. Together these questionnaires represent the extent to which the recruited participants experience pressure related to social media and body image concerns.

#### 9.3.2.2 User experience

The System Usability Scale (SUS; [1]) was used to assess user experience. This is a 10 item questionnaire that provides a usability score for the system, ranging between 1 and 100. This measure was chosen as it is brief and frequently used thus facilitating comparison with other research. The Dutch version of the SUS has previously been successfully used [12].

In addition, questions were designed specifically for the current study. Participants were asked which platform they used to engage with the training, whether they noticed the customisation of answers by the system in one of the exercises (yes/no), and how well they thought this customisation was performed ('fits very bad' to 'fits very well'). Participants were also asked to provide suggestions regarding additional useful functionalities. Finally, participants reported on how they had heard about the study, as different recruitment means were used, and whether they would like to receive updates.

[3] https://www.whatsapp.com/
[4] https://telegram.org/

#### 9.3.2.3 User engagement

Log data from the website was used to explore user engagement, including:

- Number of exercises completed and number of journal entries;
- Days on which the participant was using the system;
- Length of answers in exercises/journal.

#### 9.3.2.4 Structured interview

The structured exit interview was designed to provide information regarding both the user experience as well as the engagement of participants with the training and the study, and included question focused on:

- The user experience with the training and learning outcomes;
- Time investment and setup of the training;
- User friendliness and experience with the components of the training;
- Other remarks of the participants.

A list of open-ended questions was prepared for each topic. However, during the interviews additional questions might have been asked as well. All interviews were transcribed, from which keywords and themes were derived for the different questions. The keywords from all participants are then clustered per topic and are described in a qualitative manner. We also reported the numbers of participants who made certain statements, however this does not mean that other participants disagreed, but did not mention it explicitly.

### 9.3.3 Procedure

Recruitment was conducted via social media, through the personal accounts of the researchers as well as via advertisements. The advertisements targeted individuals of all genders experiencing body image issues and perceiving pressure from social media. Inclusion was based on age (initially 18-24 years, later extended to 30), not currently receiving therapy, and identifying as a social media user. The study was approved by the ethical committee of the Computer Science Department of the Vrije Universiteit Amsterdam.

## 9.4 Results

### 9.4.1 Participants and demographics

Figure 9.2 shows the flow of participants in the different stages of the study. Of the included participants, 21 were invited to use the system for 6 weeks and 3 could only use it for 2 weeks. All the 2-week users were interviewed, as well as 3 participants that used it for 6 weeks.

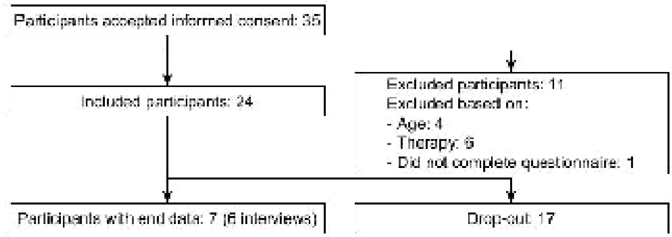



**Figure 9.2:** Participant flow

All participants were female, with a mean age of 22.5 (SD = 2.04). The majority of the participants reporting being currently enrolled in university education or having received a university degree (37.5% masters, 37.5% bachelors). A small number reported applied university education (20.8%) and vocational education (4.2%). Most participants spend more than 30 minutes per day on social media (87.5%), some participants spent more than 3 hours per day on social media (8.3%) and even fewer participants spend less than 30 minutes on social media (4.2%). During the interview half of the participants indicated knowing about and using self-compassion or some elements of it.

On average, the included participants reported self-compassion scores of 3.32 (SD = 0.78, range 1 - 7). They scored high on body surveillance (mean = 3.89, SD = 0. 81, range 1 - 5), and reported moderate range scores of body image flexibility (mean = 3.59, SD = 1.52, range 1 - 7). Concerning online photo sharing investment, participants score quite negative (mean = 3.913, SD = 0.63, range 1 - 5). Thus, the recruitment strategy seemed to have been successful in targeting individuals experiencing pressure related to social media and body image concerns.

For the remainder of the result section only the data from the 7 participants who completed the study and the final questionnaire are included, as well as the exit interviews of the 6 participants and the domain expert.

#### 9.4.1.1 Drop-out

As shown in Figure 9.2, the majority of the participants dropped out of the study without completing the final survey. Some of them dropped out before or during the tutorial (8 participants), others remained active for some time. Although not all participants informed us about their reason for dropping out, some did and informed us about a lack of time to work with it. One participant indicated that after some time the lack of diversity in the exercises led her to her drop out.

### 9.4.2 User experience of the training

Most participants (71%) used the website mostly on their laptop/desktop, others used their smartphone. Some participants explained in the interview that they used both, depending on the situation or moment of the day. During the interviews, participants described that the website worked well on their smartphones. None of the participants reported any structural problems with the functioning of the website, only some incidental bugs. On average, the total SUS-score at the end of the study was 76.79 (SD = 14.34, range 1 - 100). At the SUS item level, participants were divided about whether they would frequently use the website, with two agreeing, two neutral and three disagreeing. Similarly, for the question about being confident while using the website, three participants (strongly) agreed, one disagreed, but the other three were neutral. For the other topics participants' responses were more convergent. For example, all participants agreed that the various functionalities in the system were well integrated.

In the first exercise sentiment analysis was used to customise the reaction of the system. When asked, 57% of the participants answered that they indeed noticed this. Furthermore, on average, participants felt that the system's response was somewhat correct (mean = 3.86, SD = 0.90, range 1 - 5).

In terms of additional functionalities on the website, and other feedback, participants suggested being able to pause the training, more diversity in the exercises and a more colourful design. Moreover, some participants did not like the time restrictions on some of the dairies. Participants repeated and exemplified these remarks during the interviews.

During the interviews, participants elaborated more on the user experience of the website. First, participants discussed the overall experience with the training and whether they felt that they learned something. All participants described the functioning of the website as good and nice to work with. Participants expressed that they liked the training, and that it was easy to relate it to and apply in daily life (n=4). Moreover, the fact that it was hands-on (with clear applications) was evaluated as positive (n=3). The situations were recognisable and the training raised awareness of how to move towards a gentler attitude towards themselves (n=3). The interviewed expert agreed with this, stressing that this approach attracts a different type of audience than typical self-compassion training courses. The different components and situations helped participants to reflect on themselves, but also to see a broader perspective (n=2). The expert noted that the training could have included more focus on the process of learning about self-compassion, which is one of the main components in typical self-compassion training courses. Participants also liked how the training helped to integrate self-compassion in their daily routine (n=2). The e-mail reminders were viewed helpful for this (n=4). However, some participants continued to experience difficulties retrieving these from their spam folders (n=4). One participant indicated that an SMS would be a more convenient reminder. Participants also described that the training might constitute a first step in dealing with social media pressure on body image, or that the training might provide insight into the influence of social media and how cope with it (n=3).

It was noted by three participants that they needed more explanation or examples to understand how to practice self-compassion. This was also mentioned by the expert. Participants were asked for their views on how to engage users more over the study period. Participants pointed out that there was a lack of variety in the training, and in the situations used for the exercises (n=3). They mentioned losing interest after some time because of this. The expert suggested adding mindfulness exercises. Another way to create more variety might be to provide participants with a set goal each day (n=3), for example only one exercise and the diaries. Another participant suggested focusing on a single theme every week. Moreover, both participants and the expert suggested including more information at the beginning of the intervention regarding its functionalities, and helping users to identify ways in which to integrate the training into their daily routine.

One participant also described wishing for interaction opportunities with other participants. Participants suggested that some aspects of the training could be better emphasised in recruitment materials, such as the ease of use on a smartphone and the story element. Describing self-compassion as a way to learn to be nicer to yourself, was another suggestion.

The story used in the tutorial was described by the participants and the expert as clear, pleasant to read, and accessible and the duration was found to be suitable (n=5). Participants recognised themselves in the story and felt that is clearly explained the concept of self-compassion. However, they also recommended an introduction to put the tutorial in perspective. They would have liked it to return to this more during the training and suggested including different media as well (n=2). Finally, one participant suggested splitting the tutorial over multiple days. The expert, moreover, emphasised that some of the language used in the training could be more compassionate.

The exercise identified by the most participants as their favourite was the first exercise (n=4), in which participants were invited to react to a situation encountered by a friend and then by themselves. Participants noted that this contained all elements of self-compassion and provided insight into how to react to yourself in comparison to a friend. All exercises were clear to the participants (n=5), and participants appreciated the balance of learning and applying self-compassion (n=2). The expert noted that the form of the exercises was likely to be attractive to the target group. Although most participants found the situations recognisable, they also identified additional topics for inclusion, such as situations related to: work, school, relationships, family affairs or thought patterns. Participants' feedback was sought on the possibility of adapting the exercises to the topics discussed by the user in the diary. All participants liked this idea, but for different reasons. The expert suggested keeping this process transparent by asking participants explicitly whether they would like to work with familiar situations and later offering them an opportunity to change the situations.

Among the journals, there was not one clear favourite among the participants. A participant even described the journals as a single diary with different parts. They liked that it offered a possibility for reflection and tracking of multiple elements. However, one felt that the explanation of the purpose of journaling could have been clearer. All participants expressed disliking the fact that two of the journals were available to them only after 18h00. Regarding the mood journal specifically participants noted that in general they could always find an appropriate mood (n=3). However, they did find it hard to choose a single mood to summarize the whole day (n=4). The expert recognised this problem. Suggestions to change this included having multiple scales to represent the mood of a day, combining multiple moods. Participants noted that it would have been nice to add a description to their entry or to use the mood journal together with the self-compassion journal (n=4). Moreover, some described that seeing a pattern of negative moods could be upsetting.

Regarding the self-compassion journal, participants noted that the steps in the journal could have been more distinct and clearer, and the journal was sometimes perceived as onerous (n=2). The rating system was also sometimes found to be unclear. The expert noted that the rating system might also feel judgemental to users, this improved with different wording and/or symbols.

The story around the journey you were making was evaluated very positively by all participants. Choosing their destination was described as a nice feature and the journey did not feel competitive or something they could fail at. Reaching a destination faster, or having stops in the middle, was described as potentially increasing motivation (n=5). Finally, greater emphasis on earning the bonus every day was another suggestion for improvement (n=3).

### 9.4.3 Engagement with training

Table 9.1 shows statistics from the user log related to user engagement. It should be noted that participants from the 2-week pilot received more specific instructions about how frequently to use the training (at least 5 days). The table also shows that most participants interacted with all components the same number of times. However, some participants engaged less with particular exercises. On average, each participant provided one entry per day for each component. These data also reveal that participants were not active every day over the 2/6-week period, as indicated by the relatively low number of cities visited by each participant.

**Table 9.1:** Statistics from user log

| **Number of entries exercise** | | | **Number of entries journal** | | | | | |
|---|---|---|---|---|---|---|---|---|
| **#1** | **#2** | **#3** | **Mood** | **Gratitude** | **Self-compassion** | **KM traveled** | **Number of cities visited** | **Active days** |
| *6 weeks participants* | | | | | | | | |
| 13 | 13 | 13 | 10 | 9 | 9 | 1875.3 | 2 | 13 |
| 5 | 4 | 4 | 5 | 4 | 7 | 700.0 | 0 | 6 |
| 8 | 8 | 8 | 11 | 9 | 11 | 1400.2 | 2 | 11 |
| 6 | 5 | 6 | 12 | 8 | 5 | 650.2 | 1 | 10 |
| *2 weeks participants* | | | | | | | | |
| 8 | 5 | 6 | 7 | 2 | 6 | 775.3 | 1 | 10 |
| 3 | 3 | 3 | 2 | 2 | 3 | 325.3 | 1 | 4 |
| 8 | 8 | 7 | 5 | 4 | 3 | 900.3 | 1 | 7 |

Figure 9.3 shows the lengths of user input in the different elements with a character minimum. (Here data were combined from all the exercises we combined all exercises and steps, as each had the same character minimum.) Although many outliers were identified, on average, input was below 110.2 characters for the exercises. The data pertaining to

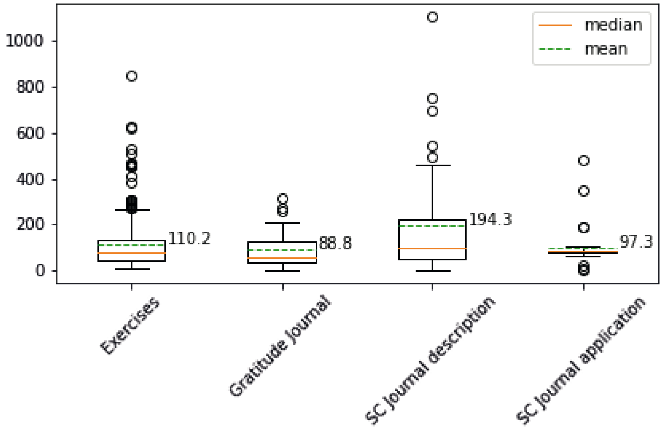



**Figure 9.3:** Boxplot of lengths (in characters) of input on different elements

the gratitude journal contained fewer outliers (Figure 9.3), although the majority of the input was still below the average length of 88.8 characters. The self-compassion (SC) journal was divided in two different boxplots: one for the description of the situation, and one for the description of how self-compassion is or could have been used. As shown in the figure, in the first step, participants provided additional details compared to the second step. In 72% of the entries in the second step the input was exactly the same as a given example. During the interviews some participants explained that they found the two steps to be overlapping.

Participants reported spending around 15 minutes per training session, although some reported spending up to half an hour, and some reported only spending 5 minutes. Participants also reported that this differed each day depending on how they were feeling or how focused they were on the training. Some reported that sometimes they were "not in the mood" to practice self-compassion (n=2). It was suggested that providing a space to include some of this contextual data would be useful. Most of the time, participants engaged in only one session per day. All participants noted that it was important to be active regularly, but opinions differed regarding the ideal frequency.

Participants were asked to describe their preferred duration of the training. All participants agreed that it should be at least 3 or 4 weeks, as interest in the training was described as declining after that amount of time. However, most participants considered 6 weeks a suitable duration, as they noted that time was needed to integrate to training into their life or recognize patterns (n=5).

Age was explored as another factor influencing engagement with the training. Overall, all participants felt that the intervention could be appropriate for slightly older individuals (up to 30 years) with some additional tailoring of the exercises. Nevertheless, finding a sustainable means of integrating the training into a daily routine was judged to be more important than tailoring to age in terms of continued engagement (n=2).

9

All participants reported that they were motivated to continue the story and progress bar with the train.

## 9.5 Conclusion and discussion

The in person nature of the majority of existing self-compassion trainings may constitute a barrier to young adults engaging with them. However, self-compassion been shown to be associated with a host of positive mental health outcomes. Therefore, the aim of this project was to develop an online, self-guided training for self-compassion for young adults. The training included three different exercises, three different journals and gamification in the form of points and narrative elements. The findings from our pilot study, conducted over a 2 or 6 weeks periods overall provide support for the format and content of the intervention. In addition, however, findings provided many avenues for refining the intervention.

Initially the training was designed to increase participants' self-compassion and prevent the detrimental effects of social media pressure and use on body image. The findings from this study suggest that participants saw a need for such an intervention, but viewed the main contributions as more broad. Participants described how in their view training contributed to learning more about being nicer to themselves across a broad range of situations. As this was not the initial goal, the situations upon which the exercises and the story line were based mainly focused on social media and body image related topics. Participants described a need for more variety in the exercises and situations, and offered suggestions. Consistent with this, the data from the self-compassion journals also revealed that participants were interested in thinking about a wide variety of topics, and more rarely social media pressure. Although broader than the original target, these findings support the usefulness of the training for increasing self-compassion in general. As self-compassion has been shown to hold many benefits in terms of well-being and seemed an acceptable and engaging topic to the current young adult participants, broadening the intervention for the larger trial seems a valuable direction.

The website was evaluated positively, as indicated by responses to both the interviews and the SUS questionnaire, although design improvements were suggested. Moreover, the website worked well both on smartphones as well as on laptops/desktops. The training itself was also evaluated positively. In particular, the practical nature of the training was seen as one of its strengths by both participants and the expert. Suggested improvements included additional explanation of the rationale underpinning the training as well as more examples related to self-compassion. In addition, increasing the amount of process guidance provided to participants was identified as a valuable improvement. Currently, the training mainly focused on providing users with theory related to self-compassion and application exercises. Providing more process guidance would include additionally focusing on participants' feelings and thoughts related to the process of engaging in the training and learning self-compassion.

Participants were very enthusiastic about the story that was built around the training and found it easy to relate to it. This was identified as another important strength of the intervention. The interview findings also resulted in suggestions for changes to the implementation of the mood journal. Additional avenues for improvement included the time restrictions placed on certain functionalities. Finally, the sentiment analysis represents another area for further refinement.

Regarding user engagement, the average time of engagement was relatively short (around 15 minutes) on a regular basis, although with some variability. Integrating the training into daily routine emerged as an important aspect to facilitate in the larger trial. On the other hand, findings highlighted that participants only interacted with the training a few times during the 6 weeks, despite being supportive of the overall length of the intervention. Moving forward the tension between maintaining the focus on self-compassion, and yet broadening the scope of the interventions in terms of topics will be an important direction.

In conclusion, the findings from this pilot study showed that a practical and story-driven approach to a self-compassion training was appealing and motivating for young adults. The intervention was originally designed to specifically target self-compassion as related to social media pressure around body image. Taken together, the current findings suggest that broadening the interventions as well as providing more theoretical rationale and support for the different active components would likely increase the acceptability, and engagement with the intervention and therefore, ultimately effective.

## Bibliography


[1] Brooke, J.: Sus: a “quick and dirty" usability scale. Usability evaluation in industry **189** (1996)

[2] Ferrari, M., Hunt, C., Harrysunker, A., Abbott, M.J., Beath, A.P., Einstein, D.A.: Self-compassion interventions and psychosocial outcomes: A meta-analysis of rcts. Mindfulness **10**(8), 1455–1473 (2019). https://doi.org/10.1007/s12671-019-01134-6

[3] Fleming, T.M., Bavin, L., Stasiak, K., Hermansson-Webb, E., Merry, S.N., Cheek, C., Lucassen, M., Lau, H.M., Pollmuller, B., Hetrick, S.: Serious games and gamification for mental health: current status and promising directions. Frontiers in psychiatry **7**, 215 (2017). https://doi.org/10.3389/fpsyt.2016.00215

[4] Johnson, E.A., O’Brien, K.A.: Self-compassion soothes the savage ego-threat system: Effects on negative affect, shame, rumination, and depressive symptoms. Journal of Social and Clinical Psychology **32**(9), 939–963 (2013). https://doi.org/10.1521/jscp.2013.32.9.939

[5] van der Lubbe, L.M., Gerritsen, C., Klein, M.C.A., Hindriks, K.V., Rodgers, R.F.: Designing a gamified self-compassion training. Proceedings of the 22nd European GAME-ON conference (2021), to appear

[6] Marengo, D., Longobardi, C., Fabris, M., Settanni, M.: Highly-visual social media and internalizing symptoms in adolescence: The mediating role of body image concerns. Computers in Human Behavior **82**, 63–69 (2018). https://doi.org/https://doi.org/10.1016/j.chb.2018.01.003

[7] McLean, S.A., Paxton, S.J., Wertheim, E.H., Masters, J.: Photoshopping the selfie: Self photo editing and photo investment are associated with body dissatisfaction in adolescent girls. International Journal of Eating Disorders **48**(8), 1132–1140 (2015). https://doi.org/10.1002/eat.22449

[8] Michael, D.R., Chen, S.L.: Serious games: Games that educate, train, and inform. Muska & Lipman/Premier-Trade (2005)

[9] Neff, K.: Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. Self and Identity **2**(2), 85–101 (2003). https://doi.org/10.1080/15298860309032

[10] Neff, K.D., Germer, C.K.: A pilot study and randomized controlled trial of the mindful self-compassion program. Journal of clinical psychology **69**(1), 28–44 (2013). https://doi.org/10.1002/jclp.21923

[11] Neff, K.D., Kirkpatrick, K.L., Rude, S.S.: Self-compassion and adaptive psychological functioning. Journal of Research in Personality **41**(1), 139–154 (2007). https://doi.org/https://doi.org/10.1016/j.jrp.2006.03.004

[12] Pluijlaar, I.: Differences between users and their user problems: a usability test for Wehkamp.nl. Master's thesis, University of Twente (2011)

[13] Raes, F., Pommier, E., Neff, K.D., Van Gucht, D.: Construction and factorial validation of a short form of the self-compassion scale. Clinical psychology & psychotherapy **18**(3), 250–255 (2011). https://doi.org/10.1002/cpp.702

[14] Sandoz, E.K., Wilson, K.G., Merwin, R.M., Kate Kellum, K.: Assessment of body image flexibility: The body image-acceptance and action questionnaire. Journal of Contextual Behavioral Science **2**(1), 39–48 (2013). https://doi.org/https://doi.org/10.1016/j.jcbs.2013.03.002

[15] Seaborn, K., Fels, D.I.: Gamification in theory and action: A survey. International Journal of human-computer studies **74**, 14–31 (2015)

[16] Stice, E., Marti, C.N., Durant, S.: Risk factors for onset of eating disorders: Evidence of multiple risk pathways from an 8-year prospective study. Behaviour Research and Therapy **49**(10), 622–627 (2011). https://doi.org/10.1016/j.brat.2011.06.009

[17] van der Lubbe, L.M., Gerritsen, C., Klein, M.C.A., Hindriks, K.V.: Empowering vulnerable target groups with serious games and gamification. Entertainment Computing **38**, 100402 (2021). https://doi.org/10.1016/j.entcom.2020.100402

[18] Veldhuis, J., Alleva, J.M., Bij de Vaate, A.J., Keijer, M., Konijn, E.A.: Me, my selfie, and i: The relations between selfie behaviors, body image, self-objectification, and self-esteem in young women. Psychology of Popular Media **9**(1), 3 (2020). https://doi.org/10.1037/ppm0000206

# Chapter 10

# Using topic modelling to personalise a digital self-compassion training

This chapter is published as: **van der Lubbe, L.M.**, Groot, N., Gerritsen, C. (2022). Using Topic Modelling to Personalise a Digital Self-compassion Training. In: Lewy, H., Barkan, R. (eds) Pervasive Computing Technologies for Healthcare. PH 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 431. Springer, Cham.

**This chapter contains:**



10

## Abstract

Young adults that struggle with mental health issues experience barriers to seek help. With our online self-compassion training we try to overcome some of these barriers. To improve our training, we can personalise exercises based on topic modelling. Data from a pilot study is used to analyse and evaluate the algorithm. Overall, the algorithm has an accuracy of 54.1% for predicting the right topic. This accuracy increases to 80.4% when considering an empty prediction to be correct as well. Although this research also shows

that our data makes the task of topic modelling difficult, it does prove to be a possibility to personalise the designed training.

## 10.1 Introduction

Young adults can struggle with mental health issues due to various reasons. Mental health disorders lead to a poor quality of life and have a high contribution to the global burden of disease [11]. Although young adults are struggling with their mental health, they perceive several barriers to seek help [5]. Among these barriers are the lack of accessibility, preference for self-help instead of external help, and a lack of knowledge of what services exist.

To overcome the barriers around accessibility and self-help, an online intervention is a promising tool. It has the convenience that it is a private and flexible way of training yourself. We designed an online self-compassion training with gamification elements for young adults (Chapter 8, [7]). Self-compassion means that you are kind to yourself in difficult times, you perceive your experiences as part of the larger human experience, and that you have a mindful attitude towards difficult emotions. It has been associated with positive outcomes, such as an association with greater happiness, optimism, and positive affect [9]. Often, training courses for self-compassion are in person, in which the trainer can have interaction with the participants [3]. However, online alternatives also exist [4, 10].

The online self-compassion training we designed is self-guided, meaning there is no professional supervision or guidance during the training. While in in-person self-compassion training such personal guidance plays an important role, we have to create an automated alternative that still provides some personalization e.g. within one of the exercises (Chapter 8, [7]). During the pilot study, that we performed to assess the user-experience of the website, we discussed the option of personalising the training content with the participants (Chapter 9, [8]). By adapting the exercises of the training to topics that participants discuss in different components of the training, it becomes possible to better suit the exercises to the needs of individual users. Therefore we have developed an algorithm for topic modelling that can be used in the improved version of our online self-compassion training.

This chapter first discusses background literature on topic modelling. Next, the existing online self-compassion training is introduced and the method of the data analysis and algorithm implementation are described. Followed by the results of this analysis and the evaluation of the algorithm. Finally, we will draw conclusions on how this algorithm can be used to personalise the content of our training website, and what lessons can be learned from a technical perspective.

## 10.2 Topic modelling

Topic modelling means finding themes in unstructured documents [2]. Different approaches and algorithms for this task exist. In [6], different ways of classifying these algorithms are described. First, the classification based on used strategy: probabilistic or non-probabilistic (or algebraic models). Non-probabilistic models use a Bag-of-Words (BoW) approach. In this approach the corpus gets converted into a term document matrix and the order of terms is neglected. The probabilistic model improves such non-probabilistic models by adding the probability sense using generative model approaches. The next classification that can be made is that of supervised and unsupervised approaches. The main difference between these two approaches is the existence of labels in the training data set [1]. Supervised modelling works with predetermined output attributes (labels). The models attempt to predict and classify the predetermined attribute, and their accuracies (alongside other performance measures) is dependent on the number of correctly classified attributes. Unsupervised modelling, on the other hand, focuses on clustering without the use of target attributes. Lastly, one can distinct whether algorithms use the sequence of words during topic modelling or use the BoW approach that does not consider this [6].

## 10.3 Method

Currently, the online self-compassion training consists of three exercises, a journal consisting of three components, and a profile page (Chapter 8, [7]). Gamification is added in the form of a story that the user progresses through. This story is about a journey that you are making: a metaphor for your self-development through learning about self-compassion. The story is a way to deliver theory to the user in a recognisable context. Moreover, to progress on your journey you have to earn kilometres, which you earn by engaging in exercises and the different journals. For each finished component you progress a certain number of kilometres, and when completing all components on a day you earn a bonus. There is a maximum progression per day, as more interaction is not considered beneficial anymore.

Initially, the story and exercises use situations written around the topics of social media, body image issues, social anxiety and troubles with friends, with as main goal reducing body image related issues of young adults. With assigning topics to previous text written by users, the choice for those situations can be personalised to the individual user. During the pilot study we discussed this possibility with users (Chapter 9, [8]). They said that they are interested in such a feature. While some users would like to practise with situations that are close to them, others noted that they would like it the other way around as that would be less personal. Moreover, users also noted that it could be beneficial to learn how to generalise the application of self-compassion, and thus it would be good to also practise with more unfamiliar situations.

Thus, the goal is to create a topic modelling algorithm that can be used to determine which topics are discussed in freely written, user created, content, which can then be used to personalise the exercises of the training. As we use predetermined topics, we cannot use existing algorithms such as those mentioned in Section 10.2. With data from the pilot study we can analyse and evaluate how the algorithm performs and how it can be embedded in the training website.

### 10.3.1 Data

In the previously mentioned pilot study, 24 users worked with the website for a limited period of time (Chapter 9, [8]). All participants were female, with a mean age of 22.5 (SD = 2.04). From these 24 users, 18 users actually entered data in the website. Of these 18 users, the data entered in the exercises and journals is saved. For this study we use the data from the gratitude and self-compassion journal, and one of the exercises called ‘Practise self-compassion’. These contain texts about situations from the users’ daily lives. All data is in Dutch.

In the gratitude journal users are asked to fill in something they are grateful for on that day, and characterise their answer with a short tag (max. 50 characters). Both this description as well as the tag can be used to determine the topic of the text. For the self-compassion journal users describe a situation, characterise this with a similar short tag and explain how self-compassion is used or could be used in that situation. From this journal we use the situation description and the tag. Finally, in the exercise users are asked if there is something they are struggling with and to describe this situation. Even if they are not struggling at the moment, they are asked to use a situation from their (recent) past. This description is used for the topic modelling. For the exercise no tag is available. In total, the available data contains 181 unique notes: 61 gratitude journal notes, 54 self-compassion journal notes, and 66 exercises. Not all participants contributed the same number of notes. Figure 10.1 shows how many notes each participant contributed.

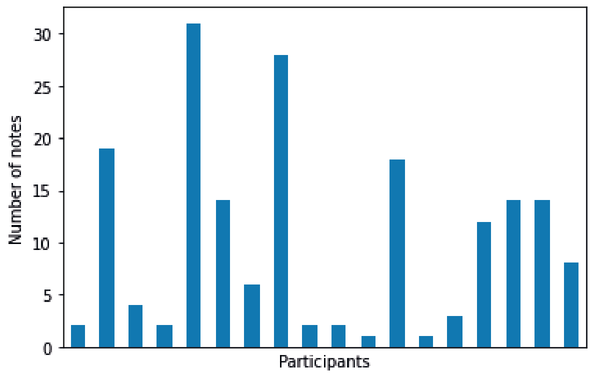



**Figure 10.1:** Number of notes per participant

### 10.3.2 Topics and data labelling

In the exercises, situations around the topics of *friendship*, *body insecurity*, *social anxiety*, and *social media* are used. The choice for these topics was related to the aim of the training. However, based on the pilot study it became clear that other topics were missing (Chapter 9, [8]). The data from the interviews of the pilot study can be used to find more suitable topics.

In addition, we analysed the data to see which topics are actually discussed by the users. To do so, the data was first labelled using the four predefined labels. If none of the labels suited the text, ‘no topic’ was noted. When the tag matched a label, this label was noted even though the text might be less clear. When multiple topics could apply, the label that the majority of the text applied to was noted. If the text was divided equally, ‘no topic’ was noted. The three researchers checked the labelling and discussed any discrepancies. The labels given by the researchers are called the gold labels. After this, we studied the notes that had no label, and decided on a new set of labels based on both the data findings as well as the results from the interviews. We changed *friendship* into *relationships*, as family or romantic relationships were also discussed. We changed *body insecurity* to *body image* as this suits better with positive message about your body. The new set of labels is: *social anxiety, social media, relationships, body image, school, job,* and *emotions*. Where emotions is a topic that covers all texts where emotions, moods or feelings have an important role. Although *social media* was not used by any of the participants, we do keep it as a topic as there is already content on the website with that topic.

### 10.3.3 Algorithm

For all predefined topics, related words were gathered using the Related Words website[1] that gives a list of words that are related to a given search term. We created our related word lists based on the terms: ‘social media’, ‘social anxiety’, ‘job’, and ‘emotions’ for the eponymous topics, and ‘body’ for the *body image* topic, ‘relationship’ and ‘friend’ for the *relationship* topic, and ‘school’ and ‘test’ for the *school* topic.

For the related word lists we first included the topic name itself. Next, we excluded words from the related word list that are clearly unrelated to our topic, such as names of people or events, drugs and disorders. For most topics we looked only at the first 50 terms. For the term ‘friend’, we only looked at words for friends or family members. For the term ‘test’, we only looked at terms related to grading in the school context. For the term ‘social media’ we included some currently popular social media platforms that are not included yet on the Related Words website. The related word lists contain 51 words for *social media*, 62 words for *body image*, 32 words for *social anxiety*, 116 words for *relationships*, 52 words for *job*, 88 words for *school*, and 51 words for *emotions*.



#### 10.3.3.1 Data preprocessing

Before analysing the data, we remove any names mentioned in the data to ensure anonimity. In most cases, the names are replaced by an X to make the text still readable for the labelling process. The texts and tags are translated to English with the Google Translate API[2]. These English texts and tags are used for the further preprocessing.

The texts are preprocessed with the help of the Natural Language Toolkit (NLTK) for Python[3]. To do this, the text is first tokenized. With this tokenization, the text is separated

[1] `https://relatedwords.org/`

[2] `https://pypi.org/project/google-trans-new/`

[3] `https://www.nltk.org/`

into words (tokens). All remaining tokens with more then two characters are saved, other tokens are removed as they are not valuable. Using the NLTK stopwords list we remove stopwords from the remaining tokens. Stopwords are words that are frequently used in human language. They are removed because they often do not add much value to a sentence. The final step is lemmatizing the words and verbs. This means that verbs are turned into their present tense and plural words are put in their singular form.

The tags in the data and the related word lists are preprocessed the same way. Due to this preprocessing, some tags might be deleted. In the preprocessing of the related words we also remove words that appear multiple times for the same topic. This can happen when multiple words included the same term, and after splitting the terms in single words these duplicate words are removed.

#### 10.3.3.2 Topic modelling

The algorithm that is developed looks at the overlap between the prepared text and tag and the prepared related words of each topic. First, it counts the number of related words that are present in the prepared text. Each overlap adds one point to the similarity score of the topic. When a word is used multiple times, each occurance is counted. Next, it is checked whether the related words overlap with the tag (if present). If this is the case, 3 points are added to the similarity score. Finally, the topic label itself is tokenized and it is checked whether the words from that tokenized label overlap with the tag. If this is the case, the similarity score is increased with 6 points.

The higher increments for an overlap with the tag are based on the fact that the tag is the shortest description of the note, so in general it has a higher likelihood of being a description of the topic of the text. When the tag overlaps with the name of the topic, this likelihood is even higher.

Once the similarity score is determined for each topic, the predicted topic is chosen. For this, we first check if the highest similarity score is higher than a threshold value. During the evaluation we will choose the right threshold value. If the highest score is above the threshold value, and this similarity score is only calculated for one topic, this topic is predicted. In other cases, the algorithm cannot be sure about the topic and thus predicts multiple topics. When evaluating the algorithm we look at the combinations of topics that are predicted and their gold labels. If we can find patterns in this, these patterns can be used to make rules about the final prediction of the algorithm. If not, the algorithm will predict ‘no topic’ when multiple topics score equally.

The algorithm needs to predict at most one topic for every text written by users for one of the data sources mentioned in Section 10.3.1. This topic will be saved for that text, but can later be manually changed by the user (choosing from the predefined set of topics).

### 10.3.4 Algorithm evaluation

To analyse the algorithm, we look at the accuracy with which it predicts the labels. To calculate this, the number of correctly predicted labels is divided by the total number of predictions. The higher the accuracy, the better. However, this accuracy is not everything. We also need to look at what goes wrong. In our application we consider it less problematic

**Table 10.1:** Numbers of notes labelled with initial labels

| | Social Anxiety | Social Media | Friendship | Body Insecurity | No Topic |
|---|---|---|---|---|---|
| **Gratitude** | 0 | 0 | 16 | 0 | 45 |
| **Journal** | 3 | 0 | 2 | 4 | 45 |
| **Exercise** | 4 | 0 | 5 | 5 | 52 |
| ***Total*** | *7* | *0* | *23* | *9* | *142* |

if the algorithm predicts 'no topic' instead of a wrong topic. Users will be able to manually add or edit a topic labelling if they want. However, if they do not correct mistakes, it is better if mistakes are avoided to prevent wrong displays of frequently used topics. Thus, we also calculate the accuracy of correct predictions and 'no topic' predictions compared to the total number of predictions. However, this accuracy cannot be used to improve the model, as that would mean that a 100% accuracy could be achieved by simply always predicting 'no topic'. Therefore we use the first accuracy to determine the best threshold for the points to be considered a labelling and to study the multiple label predictions to determine the final predictions.

## 10.4 Results

### 10.4.1 Data analysis

As shown in Table 10.1, most of the entries have no label. Furthermore, it can be seen that *social media* has not been used as a topic in any of the data. In the gratitude journal only *friendship* is discussed, which is the topic with the most positive wording. Table 10.2 shows the number of counted labels when the extended set of labels is used. Still, a quarter of the texts is not labelled.

**Table 10.2:** Numbers of notes labelled with extended labels

| | Social Anxiety | Social Media | Relationships | Body Image | Job | School | Emotions | No Topic |
|---|---|---|---|---|---|---|---|---|
| **Gratitude** | 0 | 0 | 24 | 2 | 3 | 4 | 1 | 27 |
| **Journal** | 3 | 0 | 6 | 5 | 8 | 11 | 13 | 8 |
| **Exercise** | 4 | 0 | 6 | 5 | 5 | 10 | 26 | 10 |
| ***Total*** | *7* | *0* | *36* | *12* | *16* | *25* | *40* | *45* |

Most of the notes without a label are small notes. Figure 10.2 shows that the number of texts without a tag reduces if you use a threshold for the minimum number of words in a text (without preprocessing). When using a threshold of 10 the number of unlabeled texts halves, when using a threshold of 20 it reduces to 29% of the original number of unlabeld texts and with a threshold of 30 this is only 18%. However, the other line shows that the total number of texts also reduces, at a higher rate than the unlabeled data. To include as much data as possible, we will use all notes. However, we will compare it with different thresholds to see if this effects the accuracies.

We counted the words appearing in the prepared texts of the different topics, and looked at the words that were used most frequently and more than five times. As there are no texts classified for *social media*, also no words could be found. Table 10.3 shows that most words are clearly unrelated to the topic, such as 'n't' or 'good'. However, the word 'colleague' makes sense for *job* as well as the word 'felt' for *emotions*. 'Felt' should have been changed to 'feel' in the lemmatization. However, as 'felt' is also a noun, it was not recognized as a verb. Both 'felt' and 'colleague' are added to the related word list.

**Table 10.3:** Gold labels and words that appear $>$ 5 times, * included in related word list

| **Gold label** | **Words** |
|---|---|
| Body image | good (n=6) |
| Social Anxiety | n't (n=6) |
| Relationships | grateful (n=14), good (n=8), girlfriend* (n=8), nice (n=7), n't (n=8), feel (n=8) |
| Job | today (n=7), colleague (n=7), work* (n=12), feel (n=10), job* (n=6), would (n=7), make (n=6), mistake (n=6), n't (n=7), say (n=6), message (n=6) |
| School | today (n=8), school* (n=7), lot (n=6), exam* (n=7), n't (n=6) |
| Emotions | could (n=7), feel* (n=17), felt (n=7), good (n=8), get (n=7), n't (n=6) |

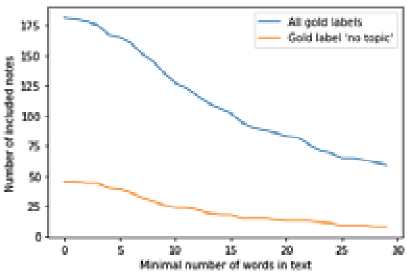



**Figure 10.2:** Number of notes when using a minimum number of words

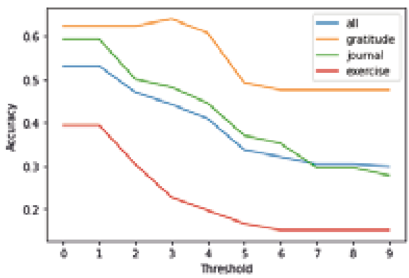



**Figure 10.3:** Accuracies for predicted labels using threshold 1-10

### 10.4.2 Results on topic modelling

First, we predict topics for each note with a points threshold of 1. If multiple topics have the highest score, we predict them all. Now we can see if there can be a pattern found in the predictions of multiple topics and their gold label. When looking at the combinations, such a pattern cannot be found. Thus, we choose that if multiple labels are predicted, the predicted label is ‘no topic’.

Next, we test thresholds 1-10 to see which threshold has the best accuracy when using the determined prediction rule for multiple labels. From Figure 10.3 it is clear that a threshold of 1 provides the best overall accuracy. Table 10.4 shows the accuracies that are predicted with this 1-point threshold. Both for the gratitude and journal notes the accuracy is higher than the average accuracy, but for the exercises this is lower. A difference between the gratitude/journal notes and the exercises is that the exercises do not have a tag. The accuracy of the 106 notes with a tag is 62.3%, the accuracy of the 75 notes without a tag[4] is 42.7%. Based on a Student’s t-test we can conclude that this difference is significant (p-value = 0.0090).

We also analysed if using a minimum of words in the original text has an effect on the accuracy, which can be found in Table 10.4 as well. Overall, this does not seem to have an effect as most of the accuracies are similar. Only for gratitude there is a big difference when using only notes that have $>$ 20 words. This category also loses 65.6% of its notes when using only $>$20 words.

**Table 10.4:** Accuracies of different components with different word thresholds. Accuracy in brackets is accuracy including ‘no topic’-predictions as correct predictions

| **Accuracy** | **All (n=181)** | **>10 words (n=127)** | **>20 words (n=83)** | **>30 words (n=55)** |
|---|---|---|---|---|
| *All* | 54.1% (80.4%) | 54.3% (79.5%) | 59.0% (77.1%) | 58.2% (74.5%) |
| *Gratitude* | 62.3% (77.0%) | 69.4% (77.8%) | 80.9% (85.7%) | 63.5% (72.7%) |
| *Journal* | 61.1% (85.2%) | 62.5% (85.0%) | 58.6% (82.8%) | 54.5% (81.8%) |
| *Exercises* | 40.9% (78.8%) | 37.2% (76.5%) | 45.4% (66.7%) | 59.1% (68.2%) |

10

Table 10.5 shows the number of notes for each topic that did not get the right prediction. The highest number of incorrect predictions are made for notes with the gold label *emotions*. However, this is also the biggest category. Relatively, the most incorrect predictions are made for gold label *social anxiety*.

As explained in Section 10.3.4, in our application it is better when the algorithm predicts ‘no topic’ instead of a wrong topic. Therefore we also calculated the accuracy when including the ‘no topic’-predictions as correct predictions, again using point threshold 1. The accuracies are shown in brackets in Table 10.4. In Table 10.5 it can be seen that still the most incorrect predictions are made for gold label *emotions*, but also for *no topic*. Relatively, *social anxiety* has the most incorrect predictions.

[4] 66 exercises notes + 9 gratitude/journal notes where the tag is removed in the preparation process

**Table 10.5:** Incorrect predictions for each gold label

| Gold label | Including 'no topic' | Excluding 'no topic' |
|---|---|---|
| **Emotions** | 27 (67.5%) | 10 (25.0%) |
| **Relationships** | 20 (55.5%) | 5 (13.9%) |
| **No topic** | 9 (20.0%) | 9 (20.0%) |
| **Body image** | 8 (66.7%) | 3 (25%) |
| **School** | 8 (32.0%) | 4 (16.0%) |
| **Social anxiety** | 6 (85.7%) | 3 (42.9%) |
| **Job** | 5 (31.2%) | 2 (12.5%) |

## 10.5 Conclusion & discussion

The goal of this chapter is to explore the possibility of using topic modelling to personalise the experience of users of our self-compassion training website. Based on the accuracy of the model it seems that to some extent it is possible to predict the topic for different texts from users. Especially when you consider a 'no topic'-prediction as a correct prediction as well. This makes sense in our application as users would be able to change or add the label manually afterwards.

For our topic modelling algorithm we choose to work with a similarity score for words and related words. Other approaches would be to use machine learning. The analysis of the data showed that on word level there are hardly any words that are characteristic for specific topics. It is thus unclear if using such approaches would make sense.

We observe that the preprocessed data loses meaning. Often the topic of a text is found by the human reader in the combination of sentences and wordings. For example a sentence like 'I have a bad headache and my shoulders are stiff due to my stress about my upcoming exam.' will be preprocessed into the words: 'bad', 'headache', 'shoulder', 'stiff', 'due', 'stress', 'upcoming', and 'exam'. Based on only this text, the algorithm will predict both the topics *school* and *body image*, as it includes the word 'shoulder' and 'exam'. However, when only looking at the prepared text it is also harder for the human reader to decide what topic this text is about. It is thus interesting to explore whether a different preprocessing could help the algorithm to perform better. However, as the data is divided over many different topics, the remaining data is limited in its size and thus it is hard to draw conclusions on this. Another note that needs to be made, is that little changes in the preprocessing could have an effct on the accuracy of the algorithm. For translating, we use an API. If something in this API changes, the texts and thus the outcomes could be different. Also, spelling errors or ambiguous words effect the outcomes of the algorithm.

Participants of the pilot were not aware of any of the topics, they were completely free when writing their notes. The data therefore is very comparable to how it will be in the actual evaluation study of our training website. This holds for the content of the data, but also for the form (length of texts, writing errors etcetera). The only difference might be that in the new version of the training the minimum length of texts will be increased based on the average lengths found in the pilot data. We could have used more data, for example from our database of situation texts, but as this data is written with the topics in mind this would not be representable for the texts of the users.

With this simple form of topic modelling we can make sure that to some degree there can be personalization in the training. We plan to include a word cloud with the topics that have been predicted by the algorithm. Moreover, we will use it to ask users if they want to practise with a situation close to them (from one of their frequently discussed topics) or something less personal (from one of the topics they do not discuss (often)). In conclusion, with the topic modelling algorithms described in this chapter, it becomes possible to personalise parts of the self-compassion training that the algorithm is developed for. With the proposed uses of this, the user-experience of the training will be increased.

## Bibliography


[1] Berry, M.W., Mohamed, A., Yap, B.W.: Supervised and unsupervised learning for data science. Springer (2019)

[2] Blei, D.M.: Probabilistic topic models. Communications of the ACM **55**(4), 77–84 (2012)

[3] Ferrari, M., Hunt, C., Harrysunker, A., Abbott, M.J., Beath, A.P., Einstein, D.A.: Self-compassion interventions and psychosocial outcomes: A meta-analysis of rcts. Mindfulness **10**(8), 1455–1473 (2019)

[4] Finlay-Jones, A., Kane, R., Rees, C.: Self-compassion online: A pilot study of an internet-based self-compassion cultivation program for psychology trainees. Journal of Clinical Psychology **73**(7), 797–816 (2017)

[5] Gulliver, A., Griffiths, K.M., Christensen, H.: Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC psychiatry **10**(1), 1–9 (2010)

[6] Kherwa, P., Bansal, P.: Topic modeling: a comprehensive review. EAI Endorsed transactions on scalable information systems **7**(24) (2020)

[7] van der Lubbe, L.M., Gerritsen, C., Klein, M.C., Hindriks, K.V., Rodgers, R.F.: Designing a gamified self-compassion training. 22st annual European GAMEON® Conference: Simulation and AI in Computer Games (2021)

[8] van der Lubbe, L.M., Gerritsen, C., Klein, M.C., Hindriks, K.V., Rodgers, R.F.: A pilot study of a gamified self-compassion training. 22st annual European GAMEON® Conference: Simulation and AI in Computer Games (2021)

[9] Neff, K.D., Rude, S.S., Kirkpatrick, K.L.: An examination of self-compassion in relation to positive psychological functioning and personality traits. Journal of research in personality **41**(4), 908–916 (2007)

[10] Talbot, F., Thériault, J., French, D.J.: Self-compassion: Evaluation of a psychoeducational website. Behavioural and cognitive psychotherapy **45**(2), 198 (2017)

[11] Ustün, T.: The global burden of mental disorders. American journal of public health **89**(9), 1315–1318 (1999)

Chapter 11

# Experiences of users with an online self-guided mental health training program

This chapter is submitted as: **van der Lubbe, L.M.**, Gerritsen, C., Klein, M.C.A., Rodgers, R.F. & Hindriks, K.V., Experiences of Users with an Online Self-Guided Mental Health Training Program, to *Journal of Healthcare Informatics Research*

**This chapter contains:**



11

## Abstract

Young adulthood is a period of high risk for the development of mental health concerns. Increasing well-being among young adults is important to prevent mental health concerns and their consequences. Self-compassion has been identified as a modifiable trait with the potential to protect against mental health concerns. An online self-guided mental health training program using gamification was developed and the user experience was evaluated in a 6 weeks experimental design. During this period, 294 participants were allocated to use the online training program via a website. User experience was assessed via self-report questionnaires , interaction data for the training program were also collected. Results showed that those who completed the intervention (N=47) visited the website on average 3.2 days a week, with a mean of 45.8 interactions during the 6-weeks. The first hypothesis that participants would report positive user experiences of the online training was supported by a System Usability Scale [4] score of 79.1 (out of 100) at the end-point. The second hypothesis regarding positive engagement with story elements of the training was also supported with scores of 4.1 (out of 5) in the evaluation of the story at the end-point.

This study found the online self-compassion intervention for youth to be acceptable, although some features seem preferred by users as compared to others. Gamification in the form of a guiding story and a reward structure seemed to be a promising element for successfully motivating participants and serving as a guiding metaphor for self-compassion.


## 11.1 Introduction

Young adulthood is a critical developmental period. Experiences during these years have long-lasting implications on many aspects of adult life [13]. It is also a period of high risk for the development of mental health concerns. Almost half of the youth who will experience mental health concerns have started to develop symptoms by age 14, and among many such concerns may remain untreated into adulthood [21]. The burden of disease of mental health disorders is high and associated with high mobility and impairment as well as negative impact on the immediate interpersonal environment [12]. For example, depression among college students can cause affective impairment or even academic impairment for moderate-to-severe levels of depression [21].

Given the course and consequences of mental health disordered among youth, their treatment and prevention are important areas of focus. However, many barriers to help-seeking exist among young adults. It has been suggested that young adults have a preference for self-help, find existing forms of help to lack accessibility and may experience difficulties with trusting the source of help [6]. Online interventions may help to overcome these barriers when they offer self-help because they are accessible as young adults are often experienced with using digital technologies, and there is no need to trust human training in fully automated interventions.

To increase the resilience of young adults in the face of difficulties, such as stress and depression, an online self-compassion training was designed and tested in a pilot study [19, 20]. Self-compassion is an attitude of observant and non-judgmental self-kindness, and can simply be explained as compassion turned inward [14]. It has been shown that higher self-compassion related to positive well-being and mental health [22], and associated with psychopathology [10, 11].

Previously, an online self-guided self-compassion training was designed for young adults, and tested in a pilot study [19, 20]. In this training, gamification was added to engage participants in the training and to explain self-compassion and its uses in an attractive way. This article describes a 6-weeks evaluation study of the online self-compassion training and presents findings related to user experience. This work explores how participants interacted with the system and the different components of the training.

In Section 11.2, background literature related to self-compassion and the mental health of young adults is presented. Section 11.3 presents the online training used for this study. Section 11.4 explains the research method applied in this study. In Section 11.5, the user experience of participants is described. Lastly, Section 11.6 draws conclusions about the user experience of the system and the hypotheses.

## 11.2 Background

Self-compassion consists of three components: self-kindness, common humanity, and mindfulness [14]. Self-kindness (1) refers to being kind and understanding towards oneself instead of being self-critical. The notion of common humanity (2) refers to receiving your experiences as part of a larger human experience, and it is the opposite of being isolated. Mindfulness (3) involves a balanced view of negative emotions, opposite to over-identification with your negative emotions. Studies have shown that self-compassion is associated with many positive dimensions of psychological health and lower psychopathology [22, 10, 11].

Several in-person interventions have been shown to be successful in increasing self-compassion. The 8-week Mindful Self-Compassion (MSC) Program [15] is one of those interventions. Digital interventions have also been studied, for example, an app to support cancer patients by learning them about self-compassion [2] and the BodiMojo app for late adolescents and emerging adults to increase their self-compassion and body image [16]. Although these apps address the same topic (self-compassion), they do not use gamification elements.

Gamification can increase user motivation and user retention [1, 7, 8]. Gamification can be defined as “the intentional use of game elements for a gameful experience of non-game tasks and contexts” [18]. A review study showed that in the mental health domain, most gamified applications targeted anxiety disorders and well-being [5]. Furthermore, this review showed that a justification for applying gamification could be found in 59% of the reviewed papers. The two main themes of justifications were: promoting engagement with an intervention and enhancing an intervention’s intended effects. The most frequently used game mechanics were: levels or progress feedback, points or scoring, rewards or prizes, narrative or theme, personalization, and customization. Although gamification is sometimes applied in interventions targeting mental health, the majority of the extant research comes from other e-health domains such as chronic disease management and rehabilitation and physical activity [17]. To the best of our knowledge, no interventions targeting self-compassion using gamification have been developed or evaluated. This article specifically focuses on points and narratives as forms of gamification and their application in an intervention targeting self-compassion.

To assess the global usability of a system in a reliable and low-cost way, the System Usability Scale (SUS) can be used [4]. This questionnaire includes 10 questions about the usability of a system that are rated on a Likert scale from 1 to 5. The total score of the SUS shows the overall system usability, which ranges between 1 (bad) and 100 (good). Bangor et al. defined different ranges to classify SUS scores [3], such as the acceptability of scores and adjective ratings to describe a system. According to the scales presented by them, scores of $>70$ are considered acceptable, scores 50 - 70 are considered marginally acceptable, and scores $<50$ are not acceptable. Adjective ratings can be used to describe the SUS scores of participants. The following adjective ratings are described by Bangor et al.: $>38$ is 'worst imaginable', 39 - 51 is 'poor', 52 - 72 is 'OK', 73 - 84 is 'good', 85 - 100 is 'excellent', and 100 is 'best imaginable'. With this assessment, an overall evaluation of a system can be given, and it allows comparisons between different systems.

## 11.3 Self-compassion training website

An online self-guided self-compassion training intervention for young adults was previously developed and described in earlier work [19]. The training was accessible via a website, where it was always available. The training was individual and self-guided, meaning there was no interaction component with other users or a mental health professional. The intervention consisted of theory related to self-compassion, exercises to practise self-compassion, and supportive journals. These components are mainly based on the MSC Program [15]. While in the MSC Program the exercises and meditations are instructed by a professional, in this intervention the instructions were given by the system. Given this lack of oversight, the intervention was designed to try and minimize the potential for negative effects and mainly focused on positive elements (e.g. focus on self-compassion instead of stress or worries in the included journals), and none of the content was focused on negative emotions or experiences.

In a pilot study, the website was tested by participants from the target group (young adults) [20]. Based on the findings of the pilot, the intervention was changed and improved. See Table 11.1 for an overview of the content, exercises and journals in the definitive version of the intervention.

The theory of self-compassion was delivered in the form of a story that users progressed through by actively engaging in the training. A significant change from the initial design was the addition of weekly themes. The story (called tutorial on the website) that was used to deliver self-compassion content and information about the intervention was rewritten and extended but remained in the same style. Moreover, it was divided over four weeks each with their own theme. The metaphor that is used in the story remained the same after the pilot: self-compassion is explained with the metaphor of a journey.

The final version of the intervention included three exercises that remained unchanged since the pilot (details on their design can be found in [19]), and a mindfulness exercise that was added after the pilot. This exercise included instructions to practice mindfulness in the present moment or an example situation that can be encountered in daily life (e.g. ‘sitting outside in the sun’). Users practised this exercise on their own and made notes if they wanted.

Another aspect of the MSC Program that was included after the pilot was choosing core values. Users unlocked this optional exercise after the first week of the tutorial. Users could choose up to 5 core values from a list, which were then shown on top of all pages to remind users what was important to them.

Additional changes were also made to the journals. First of all, the previously existing time restrictions to access the journals were removed. The remaining restrictions were that the gratitude journal could be completed once a day and that some exercises and journals were only available from a certain week and moment in the story. The restriction on the gratitude journal was set following the instructions in the MSC program to practise daily with gratitude. No other changes were made to the gratitude journal.

In the self-compassion journal, both the order and the content were modified. The new steps were that users first objectively describe a situation in their life, after which they indicate whether they used self-compassion or not and they were able to make notes about how self-compassion is connected to that situation. In the pilot study, this constituted a single step, but by dividing it into two steps, the distinction between the steps is more clear. In the following step of the journal, users previously had to give a rating on how they applied self-compassion in a situation. This step was changed, such that instead of the rating, the mood journal was now integrated into this step, and removed as a separate journal. Moreover, this step was made optional. Finally, users were able to save their journal entry as a favourite, which enables them to find the specific entry back easily. During the first week, only the first step (objective description) was available, to make users aware of situations that happen in their life that they might want to work on in the training without reflecting on it as they were just beginning to learn about self-compassion.

A new journal was also added: the sky journal. In this journal, users could write about anything they had on their minds. The journal was visualized as a sky and users could put a balloon, rainy or sunny cloud in the sky, representing something positive that happened, something they were struggling with, and challenges that they faced but turned into something positive. Figure 11.1 shows an example of an overview page of the sky journal. Although users could write notes along with their visual choice, this was not mandatory.

The game mechanics used in the intervention did not change based on the pilot. In addition to the narrative that is used, users were rewarded for completing exercises or journals with kilometres. With these kilometres they progress on their journey, ultimately reaching different destinations. In addition, by completing the daily goal, a user received a bonus in the form of additional kilometres. In this pilot study, this daily goal was to complete all available exercises and journals. However, following feedback from participants in the pilot, achieving the bonus with smaller daily goals created more variety in the training. Therefore, the daily goal was changed. In the first three weeks, the newly unlocked exercise of the week constituted that week’s daily goal. Moreover, once the gratitude journal was unlocked, this also became part of the daily goal. From week four on, the daily goal was a daily randomly selected exercise and the gratitude journal. Aside from creating more variety, the new daily goals reduced the time that users need to interact every day, as

**Table 11.1:** Intervention content and new exercises/journals per week

| Week | Theme and content | Exercises | Diary |
|---|---|---|---|
| 1 | **Introduction to self-compassion**<br>- Components of self-compassion<br>*Additional material:*<br>- Mindfulness and self-compassion<br>- Yin & Yang of self-compassion<br>- Self-compassion is different from self-esteem<br>- Which values are important to you? | **How to treat a friend** (based on an eponymous exercise from MSC Program): user reacts to a situation of a fictional friend, next they are asked how they would comfort themselves in the same situation. In the last step, they have to reflect on the differences.<br>**Choosing core values** (optional): users can choose up to 5 core values that they are reminded of in the header of the website | **SC journal** (describing situations): describe daily life situations and how SC is applied/could have been applied in those situations (based on session 'how is MSC going for me?' from MSC Program)<br>**Sky journal**: graphical journal to write about things that were positive, difficult or went from difficult to positive |
| 2 | **Self-compassion in daily life situations**<br>- Misconceptions about self-compassion are used to further explain self-compassion<br>*Additional material:*<br>- Self-criticism and safety<br>- Opening & closing | **Compassioned message** (based on 'compassionate letter to myself' from MSC Program): write a compassionate message to yourself, a good friend or to yourself as a good friend about a specific situation. | **SC journal** (complete version) |
| 3 | **Mindfulness**:<br>- Focus on mindfulness component of self-compassion | **Mindfulness exercise**: semi-guided mindfulness practice | **Gratitude journal** (based on informal practice MSC Program about counting blessings): daily write about something you are grateful for. |
| 4 | **Practice with self-compassion**<br>- Explains the stages of progress that participants can experience when increasing self-compassion<br>- The structure of remaining training weeks | **Practicing self-compassion** (based on 'motivate ourselves with compassion' from MSC Program): write a compassionate message to motivate yourself in a difficult situation | |


**Figure 11.1:** A screenshot of the Sky Journal overview page

there is a focus on a few components instead of asking users to complete all components daily. Finally, the bonus appeared more graphically (with a pop-up with text and confetti) to make users notice it more prominently and increase their motivation to reach it.

Based on the findings of the pilot, the focus of the training shifted from self-compassion as related to body image concerns towards improving well-being and life satisfaction in general by increasing self-compassion. As a result, the situations that were described in the narrative and exercises were broadened to reflect this more general focus. Based on the topics discussed by pilot participants, the following list of topics was created for exercise situations: social media, body image, social anxiety, relationships, job, school, and emotions. With the use of a specifically designed topic modelling algorithm [9], the topics discussed by users were detected in the gratitude and self-compassion journals and the 'practise self-compassion'-exercise. For the two other exercises, the user was able to choose to either practice with a familiar, unfamiliar or random topic. Familiar topics were determined as the (maximum) 4 topics that were discussed by users according to the topic modelling algorithm.

Another small change that was made based on the pilot, was the colour theme of the website to make it look more appealing. There are multiple themes that users could choose from on their profile page.

### 11.3.1 Hypotheses

Two hypotheses were developed regarding the user experience, based on the results of the previous pilot study with the system [20]. For these hypotheses, the participants who completed the end-point questionnaire, were considered when testing the hypotheses, since most data was available from them. However, their results were also compared to the results of participants that dropped out, to show differences between those who completed the intervention and those who did not. Since there was only limited data available from drop-outs, these results cannot be used to evaluate the hypotheses.

First, the hypothesis was that participants would be positive about the system's usability. This hypothesis was based on the experiences from the pilot study and the improvements that have been made after the pilot. To test this hypothesis, the SUS scores of participants were used. Moreover, answers from open-ended questions on the user experience were used to provide these scores with additional context. The hypothesis was accepted if the SUS scores could be considered acceptable and could be described with an adjective rating of 'OK' or better (score $>$52), according to the classification of Bangor et al. [3].

The second hypothesis was that participants would be engaged by the story element of the training. During the pilot, participants indicated that they felt engaged and motivated by the story element of the training. Moreover, the story element is the most important gamification aspect of this training program. To test this hypothesis, questions about the story element were added to the mid- and end-point questionnaires, which are used to evaluate the hypothesis. The hypothesis was accepted once all aspects for the story element are on average evaluated above neutral (3 out of 5). The measures used for both hypotheses are explained in more detail in Section 11.4.3.

## 11.4 Method

### 11.4.1 Design

This article aimed to study user experience and engagement during the 6-weeks training. The study implemented an experimental, between-groups design. Participants were randomly assigned to either the training or control condition. Participants in the training group were provided access to the online training for the duration of six weeks. Participants in the control condition could not access the online training, nor any other training provided in the study.

A repeated measure design was used: participants in both conditions completed questionnaires to assess their mental well-being at the start of the study, after 3 weeks (mid-point) and at end of the 6 weeks study (end-point). In addition, participants in the training condition also complete questionnaires about the user experience at mid-point and end-point.

### 11.4.2 Participants

Figure 11.2 shows the participant numbers and the rates of drop-out in different phases of the experiment. Inclusion criteria were that participants were not currently engaged in therapy for a mental health diagnosis (according to self-report) and were aged between 18 and 30 years old. The recruitment procedure is discussed in Section 11.4.4. Table 11.2 shows the demographics of the participants included in the two conditions of the study. Moreover, Table 11.2 shows that both conditions were not statistically significantly different based on their demographics.

In the training condition, 247 (84%) of the participants dropped out. The majority of this dropped out before the mid-point. This drop-out can partly be explained by the status of

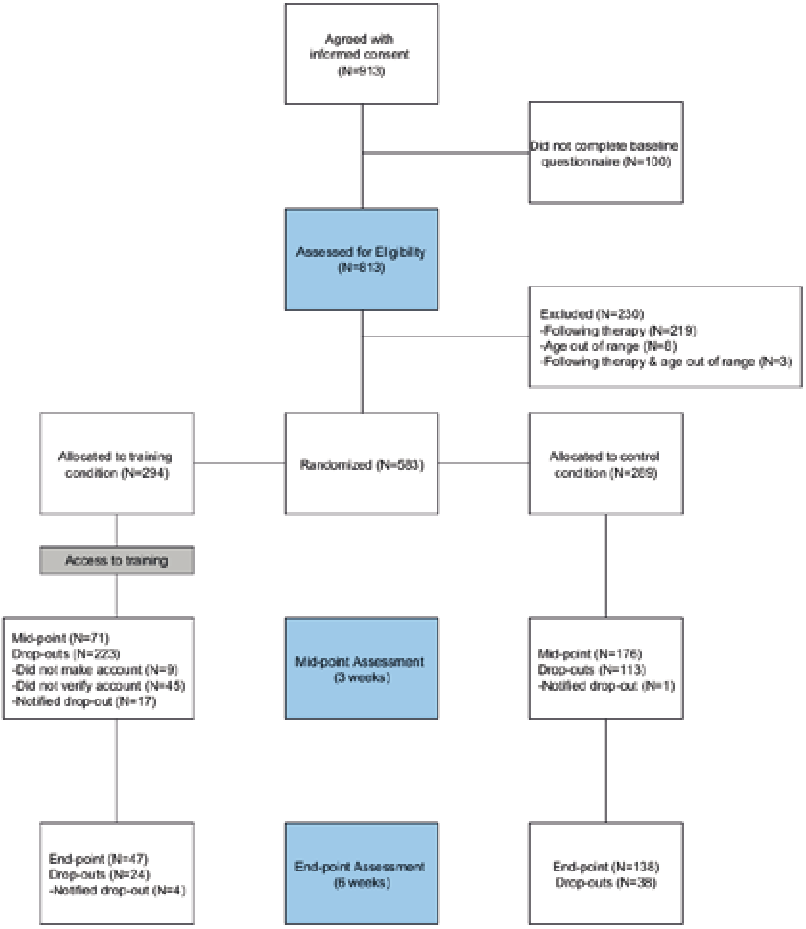



**Figure 11.2:** Participant flow

**Table 11.2:** Baseline Demographics and Sample Characteristics for the Training and Control Condition

| | Training condition (N=294) | Control condition (N=289) | Statistic |
|---|---|---|---|
| Mean age, years (SD) | 25.11 (3.381) | 25.12 (3.249) | t(581)=0.032 $p$ = 0.974 |
| Gender, n (%) | | | |
| Female | 279 (94.9) | 270 (93.4) | $\chi(3)$ = 3.676 $p$ = 0.299 |
| Male | 12 (4.1) | 16 (5.5) | |
| Other | 1 (0.3) | 3 (1.0) | |
| Wish not to share | 2 (0.7) | 0 (0.0) | |
| Highest education, n (%) | | | |
| Secundary education | 11 (3.7) | 8 (2.8) | $\chi(4)$=0.819 $p$=0.936 |
| Vocational education | 19 (6.5) | 17 (5.9) | |
| Higher professional education | 74 (25.2) | 78 (27.0) | |
| Scientific education (bachelor) | 53 (18.0) | 55 (19.0) | |
| Scientific education (master) | 137 (46.6) | 131 (45.3) | |

the user accounts of participants on the training website. Nine participants never created an account, and 45 participants did not verify their account, which together explains 21.9% of the drop-out. Moreover, 20 of the participants who dropped out completed the drop-out questionnaire (notified drop-out), which explains another 8.1% of the drop-out. One participant completed the drop-out questionnaire on the same day as (s)he received the final questionnaire, so this participant was not considered as a drop-out but did provide information about being inactive at the end of the training period. More details regarding the results of the drop-out questionnaire can be found in Section 11.5.3. In the control condition, 151 participants (55.2%) participants dropped out, of whom the majority dropped out between the start and mid-point. Participants from the training condition (N=47) that completed the intervention, completed all the questionnaires at the three time points. Of those who completed the control condition (N=138), one participant missed the mid-point questionnaire.

### 11.4.3 Materials

Participants in the training condition engaged in the training described in Section 11.3. Multiple measures were used to study different aspects of user experience and engagement with the online training. Mental health outcomes were also assessed, however, they are not reported on here.

#### 11.4.3.1 Drop-out from training

Once participants did not interact with the training for multiple days, they were asked whether they still wanted to continue. If they wished to cease their participation, they were asked to complete a short questionnaire about their reasons to quit. Various reasons for deciding to cease participation were assessed with open-ended questions and statements that were rated on a Likert scale. The questions constituting the drop-out questionnaire can be found in Appendix 11.A.

#### 11.4.3.2 User experience

Several scales were included to assess the user experience of the online training. First of all, the SUS was used to evaluate the overall experienced system usability. Moreover, a set of questions were designed for this study specifically to assess aspects of particular interest related to the user experience, see Appendix 11.B. The first aspect was the device used for the training. This question was included to be able to explore the relationship between the device used and the SUS score. Next, perceptions of the performance of the sentiment analysis used in one of the exercises was assessed. Furthermore, participants' experience of the topic detection and topic choices for the exercises was evaluated. The story and journey elements were the most important gamification elements used in the training. Therefore, participants evaluated different aspects of these elements. This included perceived effects of the story on motivation, whether the story helped to understand self-compassion and the training, and how the story was experienced and delivered. Finally, participants were invited to respond to open-ended questions about lacking features or the user experience in general.

#### 11.4.3.3 User engagement

To analyze user engagement, log data from the website were used. These data contained information about exercises and journals completed by the users and information about the user profile such as which level in the tutorial participants reached and whether they chose their core values.

With the help of these data, analyses about the number of interactions and the type of user interactions with the system were performed. Moreover, possible relationships between user engagement and other measures were tested.

### 11.4.4 Procedure



The experiment was conducted completely online. Thus, informed consent and participant responses were all collected using the software Qualtrics. Recruitment took place via paid social media advertisement and through small personal recruitment. The social media advertisement was shown to users in the target age range in the Netherlands and the Dutch-speaking part of Belgium. In all cases, participants were first directed to the online information letter and informed consent. After providing consent, participants were given an identification number and they were redirected to the first questionnaire. This questionnaire started with demographics questions, to ensure that participants met the inclusion criteria. Included participants then continued with completing the other measures about their mental health. At the end of the questionnaire, participants were randomly assigned to either the training or the control condition.

Participants that were included in the training group could immediately create an account on the website and start with the training at any moment that is convenient for them. Three weeks after completing the baseline questionnaire (control condition) or after creating your account (training condition) a second questionnaire was sent (mid-point). This questionnaire included the mental health measures, but for the training condition, it also included the user experience measures mentioned above. After another 3 weeks, all active participants were invited to complete the same questionnaire again (end-point).

Reminders for the questionnaires were sent two times during the week after the initial invitation for the questionnaires. After that, participants were considered as drop-outs, unless they completed the questionnaire at their own initiative. When it was detected that a participant stranded in the study or training, for example, the participant did complete the informed consent but did not complete the baseline questionnaire, reminders were sent as well.

Participating in the training could be seen as self-development, which was mentioned in the informed consent as one of the benefits of participating in the study. Afterwards, participants who were assigned to the control condition were granted access to the training as well, but they did not receive any more questionnaires and their data was not studied. Moreover, all participants received a debriefing information letter at the end containing information on how to continue to practise self-compassion. Participants who completed all questionnaires were entered into a lottery to win one of 10 online shopping vouchers of €25 that were provided as a financial incentive.

## 11.5 Results

User engagement and experience were examined using different statistics from the log data and the questionnaires. The hypotheses were tested in the second part of this section, where more detail on specific aspects of user engagement and experience are discussed. Lastly, this section offers more insights into the drop-out of the training condition.

### 11.5.1 User engagement and experience

#### 11.5.1.1 Activity per week

Figure 11.3 shows how many participants were active at least once during every week of the training. Active means that at least one of the exercises and journals was completed. This figure shows that in the first week, only 165 of the 240 participants with a verified account (all steps of the sign-up process are completed) were active. However, in total 175 participants were active in any week, so 10 participants only became active after the first week. Moreover, 65 participants were not active in the training at all. The training was available until participants completed the final questionnaire. Some participants interacted with the system in the 7th week, before completing the final questionnaire. This concerned

20 participants with 1 or 2 days of interactions in the 7th week. These interactions are included in week 6 in the figures in this section.

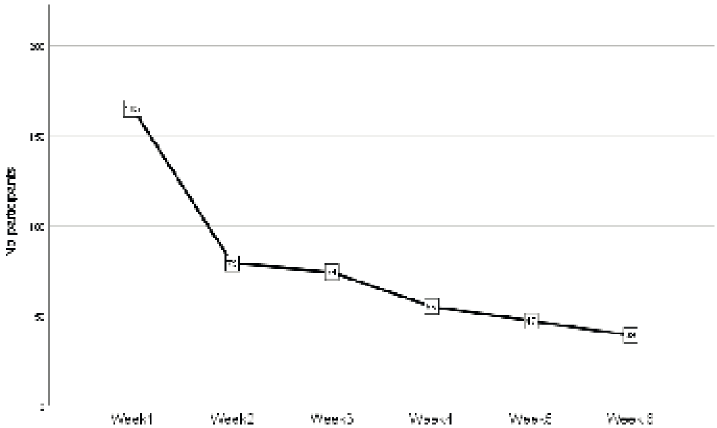


**Figure 11.3:** Number of participants active per week

Figure 11.3 only shows if participants were active at least once a week, and it should be noted this provides no information on how frequently participants interacted with the training. On average, participants (N=240) were active during 2.26 days per week (SD = 1.365). Among those who completed the intervention (N=47), the average was slightly higher: 3.17 days per week (SD = 1.521). Note that the 47 participants who were active in week 5, did not all complete the intervention. Completing the intervention was defined as completing the end-point questionnaire. It is a coincidence that the number for week 5 is also 47. Four outliers of participants that were active for more than six days a week were found. For those who did not complete the intervention and dropped out (N=193) the average was: 1.92 (SD = 1.136). For the average activity, the active weeks of the participant are used instead of the total training duration. If a participant had no activity at all, this participant was not considered in the averages.

It was also of interest to see if participants become more or less active during specific weeks. Figure 11.4 shows the average activity for each week for those who completed the study and those who dropped out from the study. The figure shows that those who completed the intervention were more active compared to drop-out participants. This difference was statistically significant in all weeks, except for week 2. The activity of those who completed the intervention was the highest in the first week but remained quite stable over time. Among those who dropped out, activity decreased sharply after the second week.



#### 11.5.1.2 Interactions per training component

This section describes the number of interactions with the website per participant. An interaction referred to completing an exercise or journal. On average, participants (N=240)

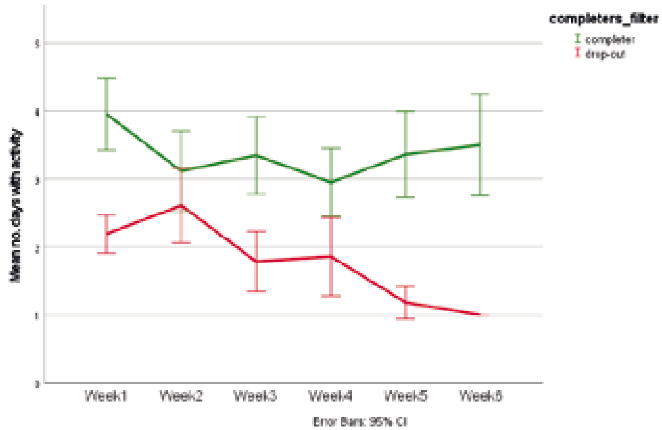



**Figure 11.4:** Mean activity per week

had a total of 13.43 interactions (SD = 21.812) with the system over the duration that they were active on the website. However, a difference can be seen between those who completed the intervention and those who dropped out. The average total number of interactions of those who completed the intervention was 45.77 (SD = 27.300), while the average of those who dropped out was 5.56 (SD = 9.771). Although these numbers cannot be compared, since the period of activity is longer for those who completed the intervention and those who dropped out, it does show that those who dropped out of the intervention on average have a low number of interactions.

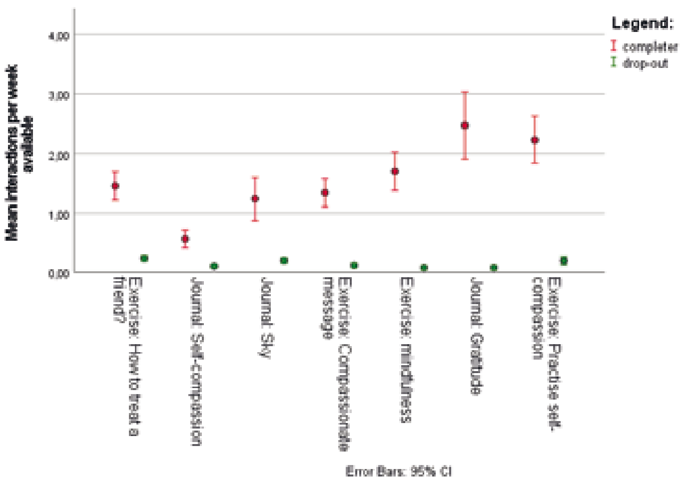



**Figure 11.5:** Interactions per program component, ordered on availability

To see which components were more popular among participants, the average number of interactions per week that a component was available can be studied (see Figure 11.5). Note that in this graph no correction was applied when participants were not active in a week. This figure shows that among those who completed the intervention the gratitude journal was the most frequently completed, followed by the 'Practise self-compassion' exercise. The mean interactions per week available of the gratitude journal was statically significantly higher compared to all other components, except for the 'Practise self-compassion' exercise. The mean interactions per week available of the 'Practise self-compassion' exercise was statically significantly higher compared to all other components, except for the gratitude journal. The least frequently completed was the self-compassion journal, which was available the whole training. The mean interactions per week available for the self-compassion journal were statically significantly lower compared to all other components.

Figure 11.6 shows the mean length of answers given by participants (+ standard deviation), for all components of the training. In the orange bars, the numbers of participants with input for that component on a given day is shown. Note that occasional interaction in the 7th week is also included in the figures. Figure 11.6a shows that the number of participants slightly decreases over time. The length of answers in the gratitude journal remains stable over time. Figure 11.6b shows the answers in the self-compassion journal. In the first week, only one part of the journal was available, after that week there was a slight increase in the length of answers. Figure 11.6b also shows that the use of the journal quickly decreases. Figure 11.6c shows the notes in the sky journal. These notes were optional and were added for 50.5% of the journal entries. Empty notes are not taken into account for this figure, both for the answer length and the participant number. Although the lengths remain quite stable, the number of participants revealed a similar pattern to the self-compassion journal.

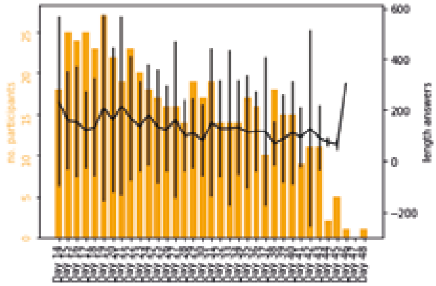
From Figures 11.6d and 11.6e, it is clear that those exercises are most used in their first week, which is different from the exercise in Figure 11.6f. This could be caused by the daily goal. In their first week, those exercises had been the daily goal, whereas the other exercise was not. For the mindfulness exercise, shown in Figure 11.6g, the notes were optional. Only 41.3% of the completed exercises contained a note. Only completed notes are included in this figure for the answer length and participant numbers. The answer length of the notes remained stable, although for the number of users a similar pattern to the other exercises is observed. Overall, all components shown in these figures show quite stable answer lengths, and no clear trends emerged.

The additional material of the first week offered participants the possibility to choose core values, which they were reminded of during the whole training on the top part of every page. Of those who completed the intervention, 78.7% chose their core values, compared to 32.1% of those who dropped out.

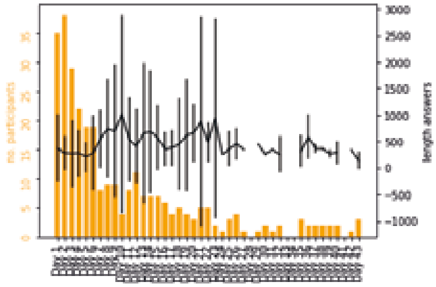


#### 11.5.1.3 Devices

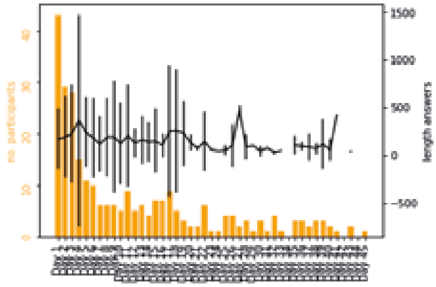
Table 11.3 shows the devices participants reported to have used the most to access the training. Smartphones were the most popular devices among those who completed the intervention, but laptops/PCs were almost equally mentioned. Tablets were not used by



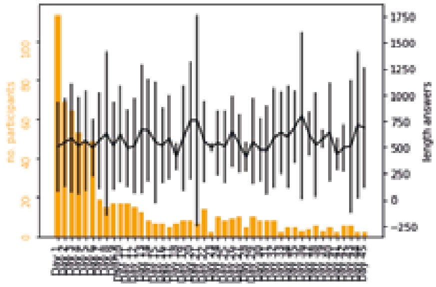

**(a)** Gratitude journal



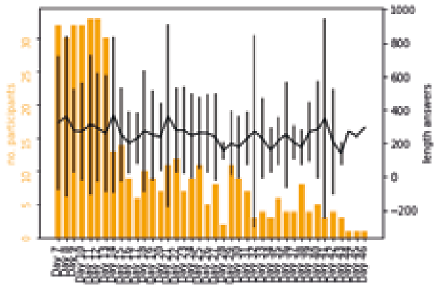

**(b)** Self-Compassion journal



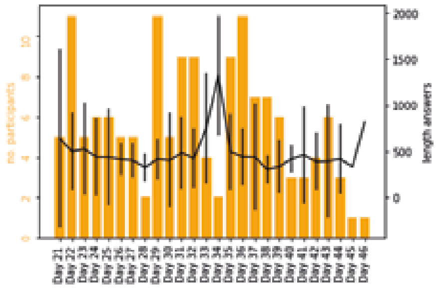

**(c)** Sky journal



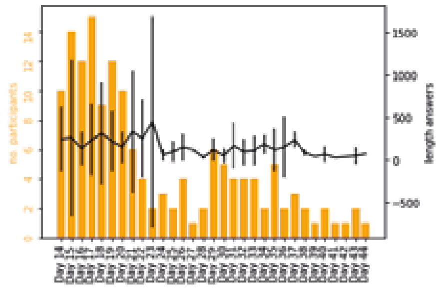

**(d)** ‘How to treat a friend?’




**(e)** ‘Compassionate message’




**(f)** ‘Practise self-compassion’




**(g)** Mindfulness exercise

**Figure 11.6:** Answer lengths per training day

those who completed the intervention. The drop-outs were using laptops/PCs slightly more often compared to those who completed the intervention. However, only a few of those who dropped out completed the mid-point questionnaire, so this view could be biased.

**Table 11.3:** Devices used by participants

| | Drop-outs mid-point (N=24) | Completers mid-point (N=47) | Completers end-point (N =47) |
|---|---|---|---|
| Laptop/PC | 13 (54.2%) | 21 (44.7%) | 21 (44.7%) |
| Smartphone | 10 (41.7%) | 26 (55.3%) | 26 (55.3%) |
| Tablet | 1 (4.2%) | 0 (0%) | 0 (0%) |

#### 11.5.1.4 Sentiment analysis

To assess how participants experienced the sentiment analysis in the 'How to treat a friend?'-exercise, two questions were used. First, participants were asked whether they noticed that the last response in the first exercise was tailored to their previous answers in the exercise. At mid-point, most of those who completed the intervention noticed this (66%), while for those who dropped out only 41.7% noticed it. Participants mostly thought that the answers by the sentiment analysis matched somewhat with their own feelings: 50% of those who dropped out chose this answer and 63.8% of those who completed the intervention. At the end-point, 78.7% of those who completed the intervention noticed the sentiment analysis, and again most of the participants thought that it somewhat matched their feelings (68.1%).

#### 11.5.1.5 Topic detection

Two questions were included in the user experience questionnaire to assess the user experience with the topic detection functionality. Those who completed the intervention rated the question about whether the topics were suitable slightly above neutral (mid-point: Mean=3.26, SD = 1.132, end-point: Mean=3.23, SD = 1.108). The mean score of those who dropped out was not significantly different ($p$ = 0.861): mid-point: Mean=3.21, SD = 0.932. Those who completed the intervention (Mean=4.26, SD = 0.706) are more positive about the question about how they liked the possibility to choose the type of situation compared to those who dropped out (Mean=3.5, SD = 0.885). This difference was significant ($p < 0.001$). Those who completed the intervention rated this question the same at mid-point and at end-point.

#### 11.5.1.6 User feedback

Besides log data and closed questions, answers from two open-ended questions offered participants additional space to describe features they felt were missing or remarks they wanted to make. All answers from participants in the training condition at mid-point and end-point are read and summarized into different categories. In total, 44 answers (mid-point with completed intervention = 18, end-point = 18, mid-point dropped out of intervention = 8) of 27 participants (22 with completed intervention, 5 dropped out of intervention) were analysed.

Most remarks were about the fact that the website was sometimes unclear to users. They did not always know where to find information, how the tutorial worked or what was expected from them. Moreover, the user-friendliness of the website was sometimes considered insufficient, and some remarks about the clarity or structure of specific components were made. A suggestion made to improve this was to add a 'frequently asked questions' page to the website. It was noted during the coding that most answers by drop-out participants were about things that were unclear or not user friendly.

Another recurring topic was the tutorial. Participants did not like that it was split up into many texts, and could sometimes not read it properly when using their smartphone. It was noted that it was sometimes hard to find information in previous tutorials. A suggestion made to improve the readability was to add page numbers to the tutorial, which could make the progress clearer. Another suggestion made was to use an interactive video or cartoon instead of the tutorial, as this participant did not like all the reading. Despite these remarks, participants also answered that they were less motivated once the tutorial ended and that it was a pity that the tutorial ended halfway through the study.

Another topic that was discussed by participants was the automated responses of the system in exercises. They did not always fit and felt like standard responses. One participant noted that (s)he would have liked to receive a response from a professional.

Some participants described difficulties remaining motivated to engage with the intervention. Some participants explained that they missed the reminder emails because these were identified as spam in their email. One participant noted that once the routine of going to school was broken due to a COVID-19 lockdown (s)he forgot to log in. However, (s)he thought about the website in situations in which the training was applicable, so (s)he did return to the website for such events.

Different suggestions of additions to the website were made. Adding more guided mindfulness exercises, supported by audio or visuals/movies was proposed by participants. Other suggestions were to add more interactive exercises, more examples of compassionate messages, and transform the training more into a game. One participant emphasised that (s)he liked the training and would like to continue to use it.

### 11.5.2 Hypotheses testing

#### 11.5.2.1 Hypothesis 1: Participants report positive user experiences of the online training

Among those who completed the intervention, the mean mid-point SUS-score was 75.69 (SD = 12.110). The mean end-point SUS-score was 79.10 (SD = 12.953). The mean SUS score for the mid-point of those who dropped out (N=24) was 69.06 (SD = 16.316). The mid-point SUS scores of those who completed the intervention and those who dropped out were not significantly different ($p$ = 0.087).

The mean SUS scores found fall in the acceptable range of scores [3]. While the score of those who dropped out can be considered as ‘OK’, the ratings of those who completed the intervention can be classified as ‘good’. These results met the criteria from Section 11.3.1; the hypothesis was therefore found to be supported.

To test whether a relationship exists between the SUS score and certain variables, such as the status of the participant (completing or dropped out), device used, age and education level, multiple regression analysis can be performed. Since 94.6% of the participants with an active account were female participants, separate analyses were not conducted among different gender. Multiple regression models were conducted to examine the variance in SUS scores at mid-point accounted for by completer status, age, education level and device used. Education level did not have a significant correlation with the SUS score. Therefore, education level was removed from the multiple regression model. The resulting model significantly predicted the mid-point SUS score, $F(3, 67) = 3.531$, $p = 0.019$. As presented in Table 11.4, higher age emerged as associated with lower SUS scores ($p = 0.032$). In addition, having completed the intervention versus having dropped out was associated with higher SUS scores ($p = 0.088$), and using a handheld (smartphone or tablet) was associated with higher SUS scores ($p = 0.124$).

**Table 11.4:** Multiple regression results for SUS mid-point

| SUS mid-point | *B* | 95% CI for *B* | | *SE B* | $\beta$ | $R^2$ | $\Delta R^2$ |
|---|---|---|---|---|---|---|---|
| | | *LL* | *UL* | | | | |
| Model | | | | | | 0.14 | 0.98* |
| Completers | 5.77 | -0.89 | 12.43 | 3.34 | 0.198 | | |
| Device | 4.93 | -1.39 | 11.26 | 3.17 | 0.178 | | |
| Age | -1.18* | -2.25 | -0.10 | 0.54 | -0.250* | | |

*Note.* Model = "Enter" method in SPSS Statistics; *B* = unstandardized regression coefficient; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; *SE B* = standard error of the coefficient; $\beta$ = standardized coefficient; $R^2$ = coefficient of determination; $\Delta R^2$ = adjusted $R^2$. *$p<0.05$

#### 11.5.2.2 Hypothesis 2: Participants have positive engagement with story elements of the training

Table 11.5 summarizes the results of the questions on the story element of the training. The results showed that those who completed the intervention evaluated the story component *understanding* the highest, followed by the component *experience of the story*, both at mid-point as well as at end-point. The lowest score at the end-point was for the component on *delivery* and at the mid-point for the component on *motivation*. Since all aspects are on average evaluated with scores $> 3$, the hypothesis was supported.

Those who completed the intervention were more positive compared to those who dropped out, who endorsed all of the aspects significantly lower as compared to those

**Table 11.5:** Evaluation story element: Mean scores per aspects (1 - 5)

| | Completers mid-point (N=47) | Drop-outs mid-point (N=24) | Completers end-point (N =47) |
|---|---|---|---|
| Motivation | 3.92 (SD = 0.891) | 3.06 (SD = 1.056) | 3.95 (SD = 0.916) |
| Understand SC and training | 4.35 (SD = 0.751) | 3.64 (SD = 0.814) | 4.38 (SD = 0.534) |
| Experience of story | 4.19 (SD = 0.824) | 3.56 (SD = 0.992) | 4.21 (SD = 0.771) |
| Delivery of theory | 4.06 (SD = 0.686) | 3.30 (SD = 0.761) | 3.88 (SD = 0.826) |
| *Total* | *4.12 (SD = 0.610)* | *3.38 (SD = 0.772)* | *4.08 (SD = 0.617)* |

who completed the intervention. This difference was significant for the subscales as well as the total score.

### 11.5.3 Drop-out

The answers to the questions from the drop-out questionnaire (N=21) were analysed and coded into five categories: technology of the training, content of the training, time investment, (mental) situation, and other. Eight participants indicated that they did not have time or that the training took more time than expected. Three participants mentioned that they chose to crease participating due to mental health concerns. Two participants mentioned other reasons such as not needing the training and forgetting about it. Eight participants in total mentioned something related to the training itself. For 3 participants the technology used was a reason for leaving the study, either because they did not like the way it worked or because they experienced problems, the other 5 participants commented about their appreciation for the content of the training, for example, that they were not motivated by it, did not like the exercises or had other means of doing the same things (for example gratitude journal).

Below are the average scores for the predefined reasons, scored 1-5 (no reason for quitting - important reason for quitting) (N=20):

- The study/website was not interesting: Mean=1.75, SD = 1.372
- I didn't like participating: Mean=2.10, SD = 1.294
- Participating costs too much time: Mean=3.15, SD = 1.585
- The website was not useful for me: Mean=2.15, SD = 1.387
- I thought it was emotionally heavy to participate: Mean=1.25, SD = 0.639
- The website did not work properly: Mean=2.05, SD = 1.605
- I didn't understand the website: Mean=2.05, SD = 1.276

The results from these questions suggest that participants often felt they lacked the time to complete the intervention, or that the intervention was too time-intensive. However, none of the other reasons seemed to be a prominent reason for quitting. These findings suggest that a combination of factors might be responsive for individuals ceasing to participate, or that the most important reason was not included in the questionnaire. Another important reason was technical problems with the website or a lack of enjoyment.

In addition, participants were asked (N=21) what could have been done differently. Ten participants said that nothing could have been changed, that it was their problem, or did not indicate anything. Two participants thought that the time investment could have been clearer from the start. The remaining 9 participants made comments about improvements to the (accessibility of the) website and the training, such as increasing the performance on mobile devices and allowing users to listen to texts instead of reading them.

#### 11.5.3.1 Training progress & drop-out

The narrative with the theory about self-compassion (called the tutorial) was divided into steps that users have to click through. It can be studied in more detail when participants dropped out, by looking at the step in which they stranded. Table 11.6 shows the number of participants and their final tutorial level. The 129 participants that dropped out in the first week include the 65 participants that never had an interaction with the system. This shows that most of the drop-outs happen in the first week of the 4-week tutorial, but there is also some drop-out among participants that completed the tutorial. Among those who completed the intervention it can be observed that 5 participants did not continue with the tutorial after the 3rd week of the training, however, they still completed all questionnaires.

**Table 11.6:** Participants and their tutorial levels

| | Drop-outs (N=193) | Completers (N=47) |
|---|---|---|
| Week 1 (level $< 32$) | 129 (66.8%) | 0 (0%) |
| Week 2 (level $> 31$ & level $< 55$) | 31 (16.1%) | 0 (0%) |
| Week 3 (level $> 54$ & level $< 66$) | 16 (8.3%) | 5 (10.6%) |
| Week 4 (level $> 65$ & level $< 75$) | 0 (0%) | 0 (0%) |
| Tutorial complete (level $> 74$) | 17 (8.8%) | 42 (89.4%) |

11

## 11.6 Discussion & conclusion

This article discussed the engagement of users with an online self-compassion training website. For those who completed the intervention (N=47), the most frequently completed activity was the gratitude journal, even though this was only unlocked in the third week. This was followed by the exercise called ‘Practise self-compassion’. The self-compassion journal was completed the least often, although it was open to users since the first week. Figure 11.6b shows that after the first week, the interaction with the journal reduces. Even when in the second week more steps were added to the journal, this did not seem to increase the interaction with the journal. On average, those who completed the intervention had 45.77 (SD = 27.300) interactions on the training website during the full study. Insights into which activities were particularly engaging to the users is important. It may be that strategically introducing activities that are known to be engaging over the course of an intervention could be a good means of increasing retention.

The first hypothesis of this study was that participants would report positive user experiences of the online training. Based on the SUS scores reported by participants, this hypothesis was supported. Although remarks about possible improvements in the functioning and clarity of the website were made by participants, the overall usability of the system is acceptable. The high acceptability of the intervention is an auspicious finding. It is likely that the previous work conducted to refine the pilot intervention with the help of user feedback increased this satisfaction. User satisfaction is a critical element for maintaining engagement and therefore these findings suggest that the current intervention would be an useful tool for promoting mental health among young adults.

The second hypothesis was that participants have positive engagement with story elements of the training. This hypothesis was also accepted. From the results, it seems that the story was mainly valued for explaining what self-compassion is and what can be expected from the exercises and journals. Based on the pilot study, it was expected that participants would mainly appreciate the motivation aspect, but this turned out to be not the most highly valued aspect. The story contributed to the understanding of participants, and it thus was an important aspect of the training. Moreover, since all aspects of the story were evaluated very positively by those who completed the intervention, this study has shown that a story is a suitable and valuable game mechanic to add to a mental health training intervention.

A large drop-out was noted in this study: only 47 participants (16.0%) of the training condition completed the study. However, half of the drop-out occurred before any interaction with the website took place. In the absence of any interaction with the website, these rates of drop-out cannot be interpreted as reflecting on the quality or the content of the training. It could be that the procedure was not clear to them, or they did experience startup problems. There is however no data to support this conclusion.

Apart from these non-starters, the data on the tutorial levels of participants showed that 64 participants did not move beyond the content of the first week. Together, the participants without a (verified) account and the participants that left the intervention during the first week make 74.1% of the drop-out. After finishing the first week, 42.3% completed the whole study. These findings regarding the timing of drop-out suggest that perhaps additional clarification of study goals and expectations might be helpful to ensure that participants fully complete the intervention. Alternatively, it might be useful to understand what expectations were not met so as to evaluate the needs of young adults who might identify such an intervention as useful but not benefit from this specific format.

The recruitment, sign-up, and instructions for the training were all automated. Moreover, a paid advertisement was used, reaching a broad audience. This could have attracted a less intrinsically motivated group of participants, which could have caused the high drop-out before interacting with the website. Once participants were engaged in the training, the drop-out was much lower. Often drop-out happened between two weeks in the tutorial. It could be that participants forgot to return after that, or that it is unclear to participants what to do and expect later. There were data to support this: from analysing the open-ended questions it became clear that comments made by participants who did not complete the intervention mostly pointed to elements that were unclear or not user friendly. This could be related to the high drop-out at the beginning of the training. Moreover, problems with automated reminders ending up in spam filters could have also contributed to the drop-out.

Regarding participants that remained active in the first weeks but dropped out later, it can be observed that they were generally less active and less positive about the training as a whole. This can be seen in Figure 11.4, where the mean weekly activity of participants who dropped out was lower compared to that of the participants who completed the study. Participants who dropped out have, on average, 5.56 (SD = 9.771) interactions in the training. Participants who dropped out evaluated the training website less positively, evaluated the story less positively on all aspects assessed and were less positive about the sentiment analysis and topic detection element. Thus, satisfaction with the intervention as a whole was associated with the likelihood of completing the intervention or not, which is consistent with previous findings and highlights the importance of user experience.

A limitation of the intervention was that most content and activities were textual. Using different media, such as audio or video, could have attracted more users and could have made the training more accessible to a wider range of users. Broader dissemination of the intervention to a wider audience would be important in future work. Thus, incorporating different types of media would be an improvement to consider.

The number of game elements used was also limited. The focus was on the story and the rewards contributing to the story. However, other game mechanics could also be added, or the current mechanics could be extended, as well as including more variety in terms of media types. For example, if more visuals were used, it would be possible to add more customization options (such as character creation) and add visual collectables (such as pets or accessories for your avatar). Although desirable, these elements were beyond the scope of the current work. However, moving forward it would be useful to incorporate them in interventions to further improve the user experience.

Another limitation was that the hypotheses were evaluated using only the data from those who completed the intervention. Since the majority of drop-out occurred before the study mid-point, only limited data are available from the participants who did not complete the intervention. If more data had been available, it would have been possible to evaluate the hypotheses by using both the data from those who completed the intervention and those who dropped out.

The second hypothesis focused on the story element, which was the most prominent game mechanic used. However, the other elements (rewards and customization) were not assessed separately. However, as they were closely related to the story element, it can be presumed that participants might also have been positive about those elements. A limitation of this study is therefore that it was not possible to determine the effect of the individual elements. However, this is often difficult, as game elements are hard to isolate and their meaning and effect can change if isolated.

The large numbers of participants who left the study between the initial registration and the end of the first week revealed that the onboarding for such a program was a challenging, yet important aspect to consider. Future work should consider tools like notifications or text message reminders to increase participant retention. Another possibility would be to integrate the training website into a program structure in which a professional could be involved. If participants feel that they are part of a larger program and that there is some supervision, this might increase the adherence. However, this could create a barrier to joining such a program, and is substantially more onerous in terms of resources. It would be interesting to study the effects of integrating the training website into a supervised program, for example in collaborations with schools, colleges or universities.

Findings related to the topic detection algorithm suggest that participants like and see potential in personalization by choosing if they want to practise with topics they have previously discussed or not. However, the feedback from participants also suggested that this aspect would benefit from further improvement. It should be noted that it could also be that participants did not notice the topics, or did not pay attention to the accuracy. Since participants did like the potential of the algorithm, future research should look into (improving) the accuracy of the algorithm.

Another promising direction for future research would be to extend and improve the narrative in the training. As participants were enthusiastic about this aspect of the training, it would be beneficial to explore this further. As mentioned, adding different media into the story could improve the user experience. Other options would be to extend the content of the story or to add more interactivity to the story, for example allowing participants to choose which topic to learn about in the next story part.

To summarize, this study found the refined online self-compassion focused intervention for youth to be acceptable, although some features seem preferred by users as compared to others. Despite the positive ratings of the intervention, retention was not high. However, the majority of participants were lost before the intervention was initiated or during the very first week suggesting that this was not related to aspects of the intervention itself. Gamification in the form of a guiding story and a reward structure seemed to be a promising element for successfully motivating participants and serving as a guiding metaphor for self-compassion. Future work is needed to best disseminate such interventions to target populations and to further improve designs to best retain and engage participants.

## 11.A Drop-out questionnaire

- I want to quit the self-compassion study. This means that I can no longer use the training website and I will no longer receive questionnaires. [Yes I quit the study, No continue with the study]
- Why do you want to quit the study? [open]
- Below you find a list of reasons to quit the study. Please indicate to what extent these reasons played a role in quitting the study. [Likert 1-5 (played no role - played an important role)]
  - The study/website was not interesting.

  - I didn't like participating.
  - Participating costs too much time.
  - The website was not useful for me.
  - I thought it was emotionally heavy to participate.
  - The website did not work properly.
  - I didn't understand the website.

- What could we have done differently so that you would not have dropped out? [open]

- Any other remarks? [open]

## 11.B User experience custom questions

All questions are translated, the original questionnaire was in Dutch.

### 11.B.1 Device used

- Which device did you use most often to visit the website? [Laptop/PC, Smartphone, Tablet]

### 11.B.2 Sentiment analysis

- In the exercise 'How would you treat a friend?' we used techniques to adapt the answers from the system to your answers. Did you notice this during practice? [Yes, No, I don't know]

- Take a look at your answers in the 'How would you treat a friend?'-exercise. Looking at the last answer of the system, do you feel that it matches your answer? [Likert scale 1-5 (do no match at all- match very well)]

### 11.B.3 Topic detection

Topics were automatically added to some exercises and journals. Based on those topics, you could choose to practise with familiar or unfamiliar situations in different exercises. To which degree do you agree with the following statements? [Likert 1-5 (totally disagree - totally agree)]

- The automatically detected topics match with the topics I discussed in the journal/exercise.

- I liked that I was able to choose which type of situation I wanted to practise with.

### 11.B.4 Story

To explain what self-compassion is about, we used a story (tutorial) in the first four weeks of the training. After those four weeks, you could still read parts of the story when you reached a destination. Indicate to what extent you agree with the following statements: [Likert 1-5 (totally disagree - totally agree)]

- The story motivated me to remain active throughout the training.
- The story helped me to understand self-compassion.
- I liked to read the story.
- Each week's story had a nice length (not too long or short).
- I liked that the story was divided over 4 weeks.
- I understood the exercises and journals better because of the story.
- The story matched with my world of experience.
- I wanted to progress in the journey and this motivated me to do exercises and fill in journals regularly.
- I reached new destinations often enough.

### 11.B.5 Other questions

- Did you miss any functionalities on the website? [No, Yes namely (open)]
- Do you have any other remarks? [open]

## Bibliography

[1] AlMarshedi, A., Wanick, V., Wills, G.B., Ranchhod, A.: Gamification and Behaviour, pp. 19–29. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-45557-0_2

[2] Austin, J., Drossaert, S., van Dijk, J., Mirkovic, J., Børøsund, E., Sanderman, R., Schroevers, M., Bohlmeijer, E.: Co-creation phases in integrating top-down and bottom-up requirements: developing a self-compassion app with cancer patients. pp. 13–14 (Jun 2020), supporting Health by Technology 2020 ; Conference date: 11-06-2020 Through 12-06-2020

[3] Bangor, A., Kortum, P., Miller, J.: Determining what individual sus scores mean: Adding an adjective rating scale. Journal of usability studies **4**(3), 114–123 (2009)

[4] Brooke, J.: Sus: a "quick and dirty" usability scale. Usability evaluation in industry **189**(3) (1996)

[5] Cheng, V.W.S., Davenport, T., Johnson, D., Vella, K., Hickie, I.B.: Gamification in apps and technologies for improving mental health and well-being: Systematic review. JMIR Ment Health **6**(6), e13717 (Jun 2019). https://doi.org/10.2196/13717

[6] Gulliver, A., Griffiths, K.M., Christensen, H.: Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC psychiatry **10**(1), 1–9 (2010)

[7] Kelders, S.M., Sommers-Spijkerman, M., Goldberg, J.: Investigating the direct impact of a gamified versus nongamified well-being intervention: An exploratory experiment. J Med Internet Res **20**(7), e247 (Jul 2018). https://doi.org/10.2196/jmir.9923

[8] Litvin, S., Saunders, R., Maier, M.A., Lüttke, S.: Gamification as an approach to improve resilience and reduce attrition in mobile mental health interventions: A randomized controlled trial. PloS one **15**(9), e0237220 (2020)

[9] van der Lubbe, L., Groot, N., Gerritsen, C.: Using topic modelling to personalise a digital self-compassion training (2021), accepted at Pervasive Health 2021, proceedings to appear

[10] MacBeth, A., Gumley, A.: Exploring compassion: A meta-analysis of the association between self-compassion and psychopathology. Clinical Psychology Review **32**(6), 545–552 (2012). https://doi.org/https://doi.org/10.1016/j.cpr.2012.06.003

[11] Marsh, I.C., Chan, S.W., MacBeth, A.: Self-compassion and psychological distress in adolescents—a meta-analysis. Mindfulness **9**(4), 1011–1027 (2018)

[12] MD, E.H., Greta Guenther RN, M., Ken Hsu MSSW, C., Kris Herman RN, M.: Depression and academic impairment in college students. Journal of American College Health **45**(2), 59–64 (1996). https://doi.org/10.1080/07448481.1996.9936863, pMID: 8908879

[13] of Medicine, I., Council, N.R.: Investing in the Health and Well-Being of Young Adults. The National Academies Press, Washington, DC (2015). https://doi.org/10.17226/18869

[14] Neff, K.: Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. Self and Identity **2**(2), 85–101 (2003). https://doi.org/10.1080/15298860309032

[15] Neff, K.D., Germer, C.K.: A pilot study and randomized controlled trial of the mindful self-compassion program. Journal of Clinical Psychology **69**(1), 28–44 (2013). https://doi.org/https://doi.org/10.1002/jclp.21923

[16] Rodgers, R.F., Donovan, E., Cousineau, T., Yates, K., McGowan, K., Cook, E., Lowy, A.S., Franko, D.L.: Bodimojo: Efficacy of a mobile-based intervention in improving body image and self-compassion among adolescents. Journal of youth and adolescence **47**(7), 1363–1372 (2018)

[17] Sardi, L., Idri, A., Fernández-Alemán, J.L.: A systematic review of gamification in e-health. Journal of Biomedical Informatics **71**, 31–48 (2017). https://doi.org/https://doi.org/10.1016/j.jbi.2017.05.011

[18] Seaborn, K., Fels, D.I.: Gamification in theory and action: A survey. International Journal of Human-Computer Studies **74**, 14–31 (2015). https://doi.org/https://doi.org/10.1016/j.ijhcs.2014.09.006

[19] van der Lubbe, L., Gerritsen, C., Klein, M., Hindriks, K., Rodgers, R.: Designing a gamified self-compassion training. In: Veloso, A., Mealha, O., Costa, L. (eds.) 22nd International Conference on Intelligent Games and Simulation, GAME-ON 2021. pp. 79–85. Eurosis, Aveiro, Portugal (2021)

[20] van der Lubbe, L., Gerritsen, C., Klein, M., Hindriks, K., Rodgers, R.: A pilot study of a gamified self-compassion training. In: Veloso, A., Mealha, O., Costa, L. (eds.) 22nd International Conference on Intelligent Games and Simulation, GAME-ON 2021. pp. 86–93. Eurosis,

Aveiro, Portugal (2021)

[21] World Health Organisation: Adolescent and young adult health (2021), https://www.who.int/news-room/fact-sheets/detail/adolescents-health-risks-and-solutions, Last accessed on 2022-2-3

[22] Zessin, U., Dickhäuser, O., Garbade, S.: The relationship between self-compassion and well-being: A meta-analysis. Applied Psychology: Health and Well-Being **7**(3), 340–364 (2015). https://doi.org/https://doi.org/10.1111/aphw.12051

# Part VI

# Discussion & Conclusion

# Chapter 12

# Discussion & Conclusion

**This chapter contains:**



## 12.1 Research contributions of the dissertation



The multidisciplinary research discussed in this dissertation illustrates how different methods can be applied to contribute to the domain of using serious games and gamification to empower vulnerable groups. Moreover, a large part of the contributions can have an impact on the wider field of research towards serious games and gamification. In this section, the main contributions of this dissertation are discussed.

Chapter 2 contributes to an understanding of the field of serious games and gamification to empower vulnerable people, and it explores the current state of the art in that domain. It offers a classification for applications within the domain, to describe their context and content. In the introductions of each of the successive parts in this dissertation, it is shown that this classification can be used to describe the applications that are developed within those parts.

Insights in the development of serious games and gamified applications is another contribution of this dissertation. Chapters 3, 4, 5, 8, and 9 focus on the background research, design process and implementation of various applications.

Part of the research in the development stage was to explore how AI techniques can contribute to personalization and adaptation within a serious game or gamified application. In the introduction, I discussed that AI offers a toolbox with various techniques. Choosing which technique is used from the toolbox is a choice that is made based on many different factors, such as the domain, target group, type of application, available data, and tasks in the application. In this dissertation, I present different efforts to use AI techniques in applications to empower vulnerable target groups. The discussed AI techniques are only a small sample of the wide variety of existing AI techniques. Moreover, each implementation of a technique is unique. For example, the algorithm I used to detect assertiveness in voices was used in a specific context: older adults and doorstep scams. The same algorithm could however also be used in a different setting, where the demands and performance could differ. Therefore, how AI techniques are used in this dissertation serves as an inspiration for future research.

Chapters 3, 7 and 10 discuss different AI components that were created to be used in the designed applications. In Chapter 3 an existing voice analysis algorithm was specifically trained to be used as a tool to measure assertiveness in the voice of a user. In Chapter 7, algorithms are described that are designed to make personalized meal recommendations. In Chapter 10, I discuss a topic modelling algorithm that can be used to detect specific topics in a text to personalize exercises to the users of the self-compassion training. Other AI techniques, such as knowledge representation and sentiment analysis, have also been used. The chapters discussing designs of applications (Chapters 3, 4, and 8) explain how knowledge representation is used in each project and which other AI components are used.

Each of the applications is evaluated within different types of studies. The work of Chapters 3, 4, 5 and 9 obtained results within different focus groups. In the studies described in Chapters 6 and 11, the evaluation is performed in the form of an experimental study. The contribution of these chapters is twofold. It contributes to understanding what the effects, strengths, and weaknesses of the different studies are, but these chapters also serve as examples of different research methods that can be used for evaluating serious games and gamified applications in different contexts.

## 12.2 Discussion of the research questions

In this dissertation, I discussed the research question:

> *How can serious games and gamification be used to empower vulnerable target groups?*

To do so, four subquestions were researched in the different parts of the dissertation. In the following sections, each of the subquestions is discussed. The main research question is discussed in the conclusion, which can be found in Section 12.5.

### 12.2.1 Discussing subquestion 1

The first research question, that is discussed in Part II, is:

> *What is the current state of the art for empowering vulnerable people with serious games and gamification?*

To address this research question, a structured literature review is performed. For this review, the scope of the domain serious gaming and gamification for empowering vulnerable people is limited to personal characteristics that lead to vulnerabilities, rather than context-specific vulnerabilities. To describe the state of the art of the domain both the context as well as the content of the found interventions are studied. This results in a taxonomy that I will describe in the next paragraphs.

The context of interventions is described by looking at the target group, the type of vulnerability risk targeted and the empowerment method of the intervention. For each of those, different categorical labels are defined, and these labels can be assigned to each intervention.

The content is described by looking at the technologies and game mechanics used, as well as the type of evaluation study that is used to evaluate the system. For the technologies and the game mechanics, different categorical labels were defined. However, in contrast to the description of the context, one intervention could have multiple (different) labels for each aspect, e.g. the labels objective, rewards and narrative as game mechanics, despite whether one or more reward mechanics are used. The study is described by one or more labels for the type of study, the duration (long; >6 months versus short; <6 months), the number of participants, and a high-level description of the found results.

Most papers describe a study with an age group as a target group, meaning that their target group is vulnerable due to their age. These groups are vulnerable to all types of vulnerability risks, but mostly for health risks often due to some (consumption) behaviour that is related to their age that is possibly a health risk, such as substance use among adolescents.

In total, most of the vulnerabilities (80%) are related to health, and for every target group label, it was found that at least some interventions addressed a health risk. For interventions with a target group of people with a physical health condition, it is found that they always focus on health risks. In most cases, the risk is coming from the physical health condition of the person, but some more indirect connections are also found. However, other risks, such as social exclusion because of a health condition, also exist among these target groups but those are not targeted. Similarly, for the target group of people with a mental health condition or a specific cultural background, all papers but one target a health risk. For the target group of people vulnerable due to their socio-economic status, health risks are not the largest category of risks, but social risks are.

The empowerment methods *training skills* and *supporting behaviour* are the most used. Skills that are trained by interventions are for example behavioural control (e.g. improving control over alcohol use) or skills relevant to prevent or treat a health condition. Supporting behaviour can have two different meanings. It can be about supporting behaviour by making physical behaviour the way of controlling the game (exergames) and thereby supporting users to make for example specific moves. However, supporting behaviour can be more indirect: the behaviour is performed without actively using the application but supported in the application, e.g. by providing monitoring options or giving rewards. Supporting behaviour is almost exclusively used to address health risks.

For the context of interventions, the review showed that most risks are very directly linked to the vulnerability of the target group, whereas more indirectly caused vulnerabilities are lacking in the current research. An example of such an indirect vulnerability can be social exclusion because of health issues, in which empowerment can help people deal with related feelings of loneliness. This is one of the research gaps that is found within the current state of the research in the domain.

Most of the technology used is mobile technology (38%). Mobile technology is the most used platform for training skills and supporting behaviour. The most used game mechanics are rewards (24%) and objectives (23%); the design of these categories is very diverse. Looking at the empowerment methods separately, rewards are the most used mechanic for each method. Especially applications for supporting behaviour use rewards frequently, namely 86% of those interventions use rewards.

Almost a quarter of the papers reviewed describe a protocol or study design but do not describe results. Most of the studies that are described in the other papers are focus groups or studies with an experimental setup with at least two conditions. Most of the studies have between 11 and 50 or 50-100 participants. Partially this is because focus groups are generally smaller compared to experimental setups. For the empowerment method *supporting behaviour*, most of the reviewed papers describe a protocol. At the same time, the most often described type of study is an experimental setup, which could explain the need for describing the study protocol alone in a separate paper. The results of the performed studies are often aimed at measuring the effect of the intervention as well as at gathering information about the user experience to improve the intervention. Long term results are often not studied and if studied effects are often not found.

Another research gap that is identified in the current state of the research in the domain is the lack of longitudinal results on the effect of an intervention. This can make it difficult to generalize results. Moreover, there is a lack of general terms or standardized forms of describing interventions. This makes it hard to compare studies.

In conclusion, on the one hand, the field of serious games and gamification for vulnerable people is a diverse domain, with applications for many different purposes and a wide variety of game or application designs. On the other hand, it was found that the type of vulnerabilities that are empowered is more uniform: they are often related to health risks and indirect vulnerabilities are often not targetted.

### 12.2.2 Discussing subquestion 2

The second research question in this dissertation, discussed in Part III is:

> *How can interactive scenarios be applied in a serious game to make older adults more verbally resilient against doorstep scams?*

To address this research question, I studied the domain of doorstep scams to develop interactive scenarios within a serious game. I closely collaborated with an organization representing the interests of older adults. Information from literature and field studies, and domain experts are used to create six interactive scenarios. Those scenarios are turn-based with multiple-choice decision moments for the player. The choices are always 'good' (assertive), 'moderate', or 'bad' (submissive). Each scenario follows a general outline: several steps (such as asking the identity card of the opponent) are present in each scenario. This makes it easier for players to remember steps that they can take during a doorstep scam to increase their chance of preventing the scam from having negative outcomes.

The interactivity of the scenarios is not only represented in the choices that the user can make. Another possible interaction that the player can have within the scenarios is through the voice analysis option. This allows users to train to use their voice assertively and influences the progression of the story.

The subquestion for this part focuses on how these interactive scenarios can be used in a serious game. As the target group consists of older adults, the design principles for this group received special attention. Ease of use of gamification declines with age [4]. Therefore, the game elements that are used are simple. For example, users receive textual feedback as well as more visual feedback with a score and a star rating. The scores are saved on a personal leaderboard. To not disturb the flow of the scenarios, the game layer is specifically built outside the scenarios. Options during the scenario, such as giving feedback after a choice is made, are explored but found to be disrupting.

Overall, participants of the final evaluation study were positive about the application. Although the application was designed with individual use in mind, participants did mention that they would like to use it in larger prevention or information sessions. It became clear that the voice analysis option did not perform as expected when used by participants, likely due to technical difficulties and performance issues when used in a live experiment. This is a point to improve in future research. Another aspect of the game that is evaluated less positive in the final questionnaire is whether participants felt personally addressed by the avatars.

In conclusion, interactive scenarios in a serious game are seen as a new way to make older adults practise their verbal resilience skills in the context of doorstep scams. With a general outline in every scenario and feedback after each practice, older adults can learn about what to say during a possible doorstep scam. With voice analysis, older adults can learn about how to say those things assertively. Game elements should mainly be used around the scenarios, to not disturb the flow of the scenario. From the evaluation of older adults, it is clear that they are positive about using the application, however, some of its performance (voice analysis, avatar quality) still need further improvements.

### 12.2.3 Discussing subquestion 3

The third research question, discussed in Part IV, of this dissertation is:

> *How to design a gamified persuasive diet tracking system as part of a diet intervention for older adults?*

To address this research question, I worked closely together with older adults (65+) and dietitians of the EU funded PROMISS project. The aim was to create a specifically designed diet tracking system that would be used within the PROMISS trial [6]: a 6-months diet trial aiming to increase the protein intake of older adults. First, the diet tracking system was designed using basic requirements and following guidelines for designing applications for older adults [1, 2]. This system consists of a tablet application and a foodbox. The tablet application uses persuasive communication. The messages that were sent to users to notify them of eating/registration moments were written according to persuasion styles. Messages could be framed using authority ('The dietician recommends...'), consensus ('A lot of people...'), or using commitment ('You've committed to...'). Using information about the personality type of the user, a different style of persuasive communication was used, following findings of previous research by Thomas et al. [8]. The prototype of this system was implemented and tested in a pilot. With this pilot, I gathered data about how participants liked working with the system, but also about what the role of the system was within the diet of the participants. An important lesson learned was that participants wanted flexibility in the app; mainly in the way that they could enter their (protein) intake: freely choosing products and using exact measures (e.g. grams, millilitres) or household measures (e.g. cups, slices). Together with the dietitians and the comments of the older adults from the pilot, the requirements of the application were refined. It was possible to create an application that facilitated different ways of entering (protein) intake, following the guidelines of the diet itself. Moreover, the application supports users during their diet intervention by suggesting products to meet their protein intake goal.

To increase the motivation of participants, the app was gamified and I added persuasive communication. The persuasive aspect was already covered in the prototype version. The gamification elements were added to the application after the pilot study, but are evaluated and improved based on an interview with an older adult. It was particularly important that the design of the gamification elements did not intervene with the diet guidelines: the rewarded behaviour needed to be following the guidelines and the reward(s) should not stimulate other behaviour. A clear example is the colour coding of the 'progression bar' for each eating moment: when a participant consumes the right amount of protein, it is coloured green, but when a participant consumes more than the advised amount of protein it turns orange, similar to the situation when you consume just under your advised amount of protein. The addition of the gamification elements had different goals: (1) increase diet compliance, (2) increase app adherence, (3) increase participants' knowledge on protein, and (4) add an element of fun. Mini-games in the form of quizzes are added to increase the knowledge of participants. The mini-games are a reward for consuming the right amount of protein, and they also serve as an element of fun and aim to stimulate diet compliance and app adherence. The reward garden that grows by playing the mini-games is meant to be fun as well but also stimulates app adherence. The profile page and achievements both stimulate app adherence and diet compliance, but do also serve as fun elements to some extent. When choosing the different game elements, I also took behaviour change techniques into account. I aimed to address multiple behaviour change techniques with the different elements of the application, as literature showed that this increases the potential of the app [9, 10]. For example, the behaviour change technique general encouragement was addressed with the achievements and coloured display of intake, and the behaviour change technique of self-monitoring of behaviour was addressed by the food diary functionality and the profile page.

With this final design of the gamified application, an evaluation study was performed during a 6-month diet trial. The results of this study show a high app adherence, however, the app did not affect the diet outcome compared to participants without a tablet nor did the gamification clearly affect the protein knowledge of participants. The experienced effectiveness, user-friendliness, notification messages and gamification were evaluated positively by participants. The foodbox was negatively evaluated. Moreover, it turned out to have less value to some participants as they wanted to store their products in the fridge or did not use the products that could be put in the foodbox.

Within the application, a limited form of adaptation to the user was performed: the suggested menu changed over time, based on previous inputs of participants. With the data gathered during the evaluation study, it was studied how machine learning algorithms can be used to give more personalized recommendations for meals and products. With the outcomes of four different algorithms, product categories and products are selected for an eating moment, keeping in mind the preferences of the user and the protein consumption required by the diet. An association rule learning algorithm is used to help the user find additional or replace products when editing a meal. The goal of integrating these algorithms in the application is to reduce the number of clicks needed to construct the meal of a participant. It is hypothesised that this improves the user experience with the application. It turns out that suggesting participants two different meals to choose and edit from, using algorithms that take into account the historical data of the user, combined with association rule learning for additional product suggestions, significantly reduces the number of clicks needed for most meal types. Since this was a data-driven evaluation, the results for the user experience could not be tested with older adults.

In conclusion, it has been found feasible to create a specifically designed application as part of a diet trial for older adults, as it was found that participants were adherent to using the application. In such a case, it is important to take into account the specific needs of the participants, but also carefully match the requirements and design of those with the guidelines from the diet. An important lesson is to balance the flexibility of the application (freedom in entering consumption) with design guidelines for older adults. Moreover, using historical data with machine learning algorithms to personalize recommendations was found to be a way to reduce the number of clicks needed to create a menu, which could further increase user-friendliness.

### 12.2.4 Discussing subquestion 4

The fourth subquestion in this dissertation, that is discussed in Part V, is formulated as:

> *How can gamification be used to develop a self-guided intervention to enhance self-compassion and increase the well-being of young adults?*

To address this research question, a prototype of a self-guided online training using a website to enhance self-compassion was developed. An existing and evaluated in-person self-compassion training program [5] was used as the basis for the content of the training program. Practises and theory from that program were translated into the design for the self-guided intervention. Because of the self-guided nature of the intervention, it was important to keep a positive tone in all exercises. Triggering more negative feelings could be a risk for participants since there is no professional supervision.

To engage participants in the intervention, gamification has been added to the intervention. The gamification is designed around the content of the intervention. The most important element is the story element, which is used to provide an accessible way for users to learn about self-compassion. Going on a journey was used as a metaphor for enhancing your self-compassion. This metaphor was used throughout the training, for example in the storyline but also for the rewards. Participants earned kilometres by completing activities of the intervention. With these kilometres, they could reach different destinations in Europe that would unlock stamps and short informative stories in their passport.

The pilot study that was performed with the prototype gave valuable insights into how participants perceived the training. Participants were very enthusiastic about the concept of the story: it was relatable and motivational. However, they also had suggestions to improve the structure of the training, especially for making it more interesting over a longer period. Examples of those suggestions were to make the story more extensive, including more focus on the process of learning about self-compassion, and a need for more examples of how to practise self-compassion. As a result of this, the content was extended and thematised per week in the final version of the intervention. Moreover, more guidance through the experience of working on your self-compassion was added to the story and other components (e.g. including information about stages of progress in the story). Changes were made to existing exercises and journals based on the feedback of participants in the pilot. For example, the mood journal is removed and tips about writing compassionately are added to the exercises. A new exercise and journal were added based on the new story content and the feedback of the participants on wanting to have a journal to write about anything they want.

The resulting final intervention was used in an experimental study of 6 weeks in which 294 participants were assigned to the training condition in which they worked with the intervention and 289 participants that were assigned to the control condition. In this dissertation, I study the user experience of the training.

Participants were positive about the website as shown by the results of the System Usability Scale and open questions about remarks. The participants who completed the study interacted with it multiple times a week during the whole study. The most popular activity is the gratitude journal, which is only unlocked in the third week of the program. This journal is only open once a day, so the popularity cannot be due to multiple interactions per day. The main gamification element was the story, this element is evaluated positively. Participants mainly say that it helps them to understand self-compassion and the training. Moreover, participants said that it was a motivation to continue.

Although these results provide insights into the user experience of the self-guided intervention, no conclusions about the effectiveness of the intervention can be drawn. Mental health assessments were conducted during the study period. However, results on the effectiveness of the intervention based on these assessments were beyond the scope of this dissertation. Moreover, since there was only one version of the intervention, which included gamification, such results would not say anything about the added value of gamification, but more about the effectiveness of the intervention as a whole. Preliminary evaluations of the assessments are promising but will be evaluated in-depth in future research.

In the study, a high drop-out of 247 participants was experienced. Based on the available information it was found that a large part of this drop-out took place before interacting with the training. For example, some participants never interacted with the training website after completing the sign-up or did not verify their training account. An important finding of this study is that the risk of drop-out for such an automated enrolment and intervention is high, but once participants are engaged for more than one week, the drop-out lowers. Based on the available data from the mid-point questionnaire, it is found that drop-out participants are in general less engaged with and less positive about the intervention.

To personalize the experience with the training, two main AI techniques were applied. Sentiment analysis was used in one of the exercises to adapt the response of the system to previous answers by the user. From the evaluation of the experience of participants, it became clear that participants felt that this adaption was only partly suitable, or it was sometimes not even noted by participants. Another AI technique was topic modelling. A specific algorithm was created that could detect predefined topics in specific texts of users, using similarity scores with related word lists. Based on these topics, situations can be chosen in other exercises, to allow users to practise with familiar or unfamiliar situations. From the evaluation of participants, it became clear that they like the idea, but that the detected topics were not always suitable in their opinion. The topic algorithm was built using the data from the pilot study. A list of topics was constructed based on the results from analyzing texts written by users of the training website and interviews with participants. The algorithm was tested on the texts written by users from the pilot study and had acceptable yet limited accuracy.

In conclusion, the feedback of participants showed that there is room to further improve the system, but the current system did create an appealing way to learn about self-compassion for young adults. Using a story offers an attractive way to provide information about self-compassion to participants. With AI techniques such as sentiment analysis and topic detection, more personalization can be added to the experience.



## 12.3 Ethical aspects

Ethical aspects are important to keep in mind when designing applications that will be used by people. In the work in this dissertation, this is especially important since the target groups are vulnerable. For each project, ethical aspects are carefully considered and approval from the ethical committee was obtained when needed. In the chapters, I have briefly mentioned ethical aspects, such as the safety of participants in the self-compassion training program, as mentioned in Chapter 11.

In the introduction of this dissertation, the eFRIEND framework [3] was mentioned as a framework that can be used to assess the ethical aspects of intelligent environments, or in this case serious games and gamified applications. Each of the following subsections describes a principle from the framework. I will reflect on how these principles need to be addressed for serious games and gamified applications for vulnerable target groups, and illustrate how these principles were addressed in the different parts of this dissertation.

### 12.3.1 Non-maleficence and beneficence

Non-maleficence means not developing a system that will cause harm to primary users. Beneficence means that the system is meant to create social benefits for the users.

The essence of the field discussed in this dissertation is to empower vulnerable groups, and thus do no harm to them. Although the overall goal of such applications is not maleficence, applications can have unwanted and negative side effects, such as rewarding incorrect execution of behaviour (or exercise) or resulting in overuse by the user [7]. Therefore, game elements must be closely related to the desired behaviour, and if needed limited to constrain users and prevent overuse. In general, adding limits to the rewards that can be earned per day (or any other period) is a way to do this, but the limits differ per context.

Gamified applications often focus on extrinsic motivation, while intrinsic motivation is sometimes undermined [7]. A result of this can be that users' health behaviour becomes dependent on gamification. It is important, yet challenging, to tackle this problem.

More verbal resilience (Part III), a protein-rich diet (Part IV), and better mental well-being (Part V), are all examples of benefits for the user. During the development, one of the important points of attention was balancing the game elements with the goal of the application to prevent undesired side effects. For Part III this meant that I chose to build the game layer outside the scenarios, but kept the scenarios close to real-world scenarios. In Part IV, I made sure that the mechanics stimulated the right number of proteins. This resulted in the fact that only the right consumption was rewarded as the best possible behaviour, and overconsumption was, similar to underconsumption, rewarded less. In Part V, I limited the number of kilometres (points) that users could earn daily, to not encourage excessive use of the system. Altogether, such measures ensure that the system is non-maleficence, both in its main goal as well as in reducing possible side effects. However, no explicit measures are taken to prevent users from becoming dependent on the gamification elements in the different projects.

Despite the fact that the main goal of the applications in this dissertation is not to harm participants, it can be difficult to ensure this. Especially when dealing with delicate topics such as mental health (Part V). Since this mental health intervention did not include professional guidance, it is possible that users of the intervention do not receive help when needed. Especially since the system is not capable of recognizing this. To reduce the risks of this, participants were only included if they did not follow therapy for a mental health diagnosis and they were reminded of where they could reach out for professional help if needed. Moreover, the wordings used in the exercises and journal were framed positively (e.g. focus on self-compassion instead of stress or worries in the included journals) and written in close collaboration with a psychologist. However, questioning whether participants followed therapy was based on self-report, thus it could still be that participants with existing mental health problems were included.

### 12.3.2 User-centred

User-centred means that the views of various stakeholders, mainly users, should be considered throughout all stages of the project.

This principle is often addressed in research towards serious games and gamified applications since it is a good design practice to take into account the end-users, which are the most important stakeholders. In some cases, other stakeholders also need to be involved, for example, domain experts or partners that will distribute or use the applications together with the target group.

In the design process of all the applications described in this dissertation, the user plays an important role. Other stakeholders are also involved in different stages of the development and evaluation. Both older adults and domain experts were consulted in different stages of the development of the serious game to train against doorstep scams (Chapter 3). Older adults were already involved in an early stage of the design process of the PROMISS application in a pilot study (Chapter 4). Furthermore, the dietitians involved in the project were also actively involved in the redesign process (Chapter 5). For the self-compassion training, I took into account the wishes and demands of possible end-users, by organizing a pilot study and doing different interviews (Chapter 9). For this project, I also collaborated with a domain expert.

Although the approach was user-centric, it is difficult to make an end-product that is fully user-centric. Compromises need to be made between the requirements of different stakeholders or participants with conflicting wishes. For example, the work of Part IV showed that while some participants liked the gamification elements, others did not. Therefore, for some participants the current setup can be considered less user-centric, as it includes features they do not like and/or use.

In the evaluation of all the projects, I also address the user experience. Because of this, I know what could potentially be improved, and how the user experience might have influenced the effect and overall experience with the systems.

### 12.3.3 Multiple user groups

For this principle, the preferences and requirements of multiple user groups must be considered.

Applications to empower vulnerable target groups often have one very specific user group. However, multiple user groups may exist, for example when the same system can be used by parents and children or professionals and patients. In such cases, both groups must be equally involved in the development and evaluation phases.

The applications described in this dissertation all have one specific targeted user group. Therefore, I did not take into account the wishes of multiple user groups, as those are not relevant for this work.

### 12.3.4 Privacy

The privacy principle is about the control that the user has over its data and that users have sufficient information about privacy risks. Moreover, users should be able to convey their privacy requests and preferences.

For serious games and gamified applications, this is a very important topic, as it is often needed to gather data to ensure that the application is working and to study its use and effects for scientific purposes. Since the target groups are often vulnerable, this means that their data is often sensitive. Reducing data collection to the necessary minimum is, therefore, a good practice. It should also be considered that some vulnerable groups might additionally help with understanding information about privacy. It is therefore important that researchers pay attention to how they communicate information about privacy risks and adjust this to the target audience.

The applications developed in this dissertation are standalone systems. Data is only collected within the system, and no additional data (like cookies or location) are collected. Moreover, all of the systems are focused on single-user usage. Users are not able to share data with others in the application. Data is however collected for research purposes or to make the system function correctly. Participants are informed about data collection via an informed consent form.

However, if the systems would be used outside of a research setup, additional privacy measures should be taken. Privacy statements need to be made to replace the informed consent forms to make users aware of what their data is used for, e.g. users need to be made aware when their data is used by third parties. For example, diet tracking data from the application discussed in Part IV could be used for targeted advertising by grocery stores. Aside from knowing about such issues, users should also have options to manage their privacy settings, for example disabling sharing with third parties. Such settings were not prepared yet.

### 12.3.5 Data protection

Systems should comply with data protection legislation, but users should also be able to have control over their data and how it is used.

As argued already in the previous principle, it is always important that the data of users is handled with care since it might contain sensitive information. What is needed to protect data depends on the type of data and the type of application. Data protection is important for all systems that collect data, despite whether they target vulnerable groups. However, as argued in Section 12.3.4, vulnerable groups might share more sensitive information which makes data protection even more important. It is good to consider whether a system needs to be an online or offline system to achieve its goal. Risks for data protection can be reduced if a system can be used offline and/or on dedicated and protected devices.

Users should also be aware of what and why data is collected. In scientific research, the informed consent that users have to sign is how this is communicated. However, users can also be reminded of this within the system, for example by clearly asking permission to collect certain (sensory) data or by ensuring that they can see which data is collected. Many applications give users insight into their own data, for example by showing the number of steps walked or the number of exercises performed. This makes users aware of what data is collected. Again, it is important to consider that some vulnerable groups might need additional help to understand informed consent, and the use of language should always be appropriate for the target group. Finally, users should have the possibility to withdraw their data at any moment.

For offline systems, such as those described in Part III and IV, the data is less accessible. Especially because the systems were provided to participants on dedicated tablets, which were not connected to the internet and returned to the researchers after the study. Thus, outside threats such as hackers were no concern in this study design. Participants were aware that the researchers could access the data, but not remotely. For Part V, which is an online system, this is an important aspect. Different safety measures are in place. For example, the website uses a secure connection protocol (HTTPS), which is a safer way of sharing personal details compared to the non-secure, standard, protocol (HTTP).

In all cases, data is stored for research purposes. Data plans are made to ensure that data is stored safely. The informed consent that participants have to sign, also explains how data protection is handled. Moreover, through informed consent, participants were informed that they could withdraw their data at any moment. Although this is sufficient for research studies, other data protection measures are needed once systems are used on a larger scale. Following the remarks in Section 12.3.4, users should be made aware how, where and why their data is stored. Legislation, such as the GDPR[1], should be followed.

### 12.3.6 Security

Systems should have appropriate and adequate security, mainly for the data that is being collected and used.

In Section 12.3.5, I have already explained that data should be carefully handled, especially when it concerns sensitive data. Security is a general topic for all (online) systems. It is important to follow the latest insights in good security practices to ensure the safety of data in the system.

In the previous section, the importance of proper data protection and data protection measures are explained. With such measures, attention is given to the security of the system. Moreover, once studies have ended, data is moved to secure cloud storage for researchers.

Once systems are used outside of a research setting, or at a larger scale, security becomes even more important. Addition measures, such as multi-factor authentication and encryption should then be considered.

[1] https://gdpr-info.eu/

### 12.3.7 Autonomy

Users should be able to change levels of autonomy in a system. For intelligent environments, this means that users can take over tasks from the agents in the system.

For serious games and gamified applications, it is important that users have the freedom to stop using it and that they are not forced to continue to use it. Moreover, within the applications, users can get autonomy when they can change certain goals or features. However, it is sometimes challenging to combine this principle with the goal of the application. For example, if you want to stimulate a user towards a certain step count goal, it might not be beneficial if the user can adjust this step count goal. However, users always have the autonomy to deviate from the advised behaviour of the system. More indirectly, gamification causes users to lose some autonomy, due to the extrinsic motivation caused by the game elements.

Similarly, freedom and autonomy are offered to users of the systems described in this dissertation. They always have the freedom to stop using the system and are not forced to use it in any way. Within the applications the freedom is sometimes limited, for example, because certain exercises are still locked, this is related to the goal of the application. However, it can be said that users lose some autonomy over their own behaviour by using gamification; their behaviour is influenced by the game elements (extrinsic motivation).

### 12.3.8 Transparency

Transparency means that users know how the system works and what it aims to influence.

Previously, it was discussed that users need to be aware of the data that is collected from them. In addition to this, users need to be aware of the goal of applications. It is often clear to users what the goal of a serious game or gamified application is, as that behaviour or goal often has a prominent role in the application. In a research setting, informed consent is used to create transparency about the purpose of the research. In addition to clearly communicating the purpose, expectation management can also be relevant for transparency. When working with automated systems, users need to be aware that they are receiving automated feedback instead of expecting personal feedback from a professional. In addition, the system itself should be transparent as much as possible. For example, communicating why certain decisions are made and giving direct feedback.

In the informed consent forms of all my projects, participants are informed about the system they are using and the study they are participating in. Moreover, informed consents also play a role in expectation management. For example, participants of the self-compassion study (Part V) were made aware that they were using a fully automated system, and that no human professional was involved in the training. However, signing an informed consent form does not mean that participants correctly understood how the system works and what the consequences are of using the system.

The systems also offer (some) transparency. However, sometimes giving transparency during an interaction can affect the outcome. For example, the topic detection used in the self-compassion training (Chapter 10) does not explain what it is based on. However, when designing such features a trade-off needs to be made between transparency, user friendliness, and the goal you want to achieve.

### 12.3.9 Equality, dignity, inclusiveness

For this principle, the system must be accessible and affordable to the primary user group.

When dealing with vulnerable target groups, there are different aspects of accessibility and affordability that need to be taken into account. Accessibility can mean two different things. First of all, you have to make sure that your target group has access to the platform or device you want to use. If additional materials are needed, these need to be affordable for the target group. Another meaning of accessibility is that the application itself can be easily used by the target group. This includes ensuring that texts and images are readable, for example by giving them the option to enlarge font sizes. It also includes that information and navigation is understandable, for example by making sure that the level of reading difficulty of texts is suitable for the readers.

Within the design of the systems in this dissertation, the focus was on the majority of the target group. When designing for older adult users, I specifically paid attention to design guidelines to facilitate this user group (Parts III and IV). However, it was not always possible to take certain minorities within those groups into account for the accessibility of the design of the website. This is mainly due to limited resources. For example, the self-compassion training website contains a lot of text (Part V), which could have been replaced with another type of media content. For some participants, this might have been a problem, for example, because of difficulties with reading. If the systems would become available for a larger audience, it would be important to improve the accessibility of the systems. Otherwise, people with specific accessibility needs would be excluded from using the systems. Other problems with accessibility could also arise when systems would be commercialised: for example, if users need to pay for a system, or need to link a system with other accounts or services; such things can limit how accessible a system is.

## 12.4 Future work

This dissertation has contributed to the field of serious games and gamification. This section describes different directions for future work, based on the work presented in the chapters of this dissertation.

A research gap that is defined in Chapter 2 is that there are no standardized ways of describing serious games/gamification designs. Some good examples of a detailed, yet structured design description were found in the review performed in Chapter 2. Moreover, throughout this dissertation, I have shown that the taxonomy from Chapter 2 can be used to describe different types of games and applications. Working on a standardized framework to describe designs, would benefit not only the specific field of serious games and gamified applications for vulnerable groups but also the more broad research domain of serious gaming. Therefore, more theoretical research in this direction is a promising future direction. If more papers use such a standardized framework, it becomes easier to compare designs. It would be interesting to see if such an approach could lead to more understanding of which game mechanics are suitable in which types of applications, considering their effect on a specific target group.

In the design and development phase, evaluation is often performed in focus groups or pilot studies. For the research purposes discussed in this dissertation, the sample sizes achieved were sufficient. Even with smaller groups, such as the focus groups in Part III, I managed to get enough data to answer the research questions that I had. Smaller group sizes were compensated with more in-depth data gathering methods, such as interviews. However, in future research, the sample sizes should be larger to study the more generalized effect of the applications.

Chapter 2 identified that longitudinal studies are often lacking. It would be valuable to perform more long term studies to see if effects or user experience remain or change. For example, for the training described in Part V, it would be interesting to see if the effects are retained over time. However, not all applications are suitable for long term use or evaluation. For example, the serious game about doorstep scams (Part III) is not designed for repeated use due to a limited number of scenarios. However, it is designed in such a way that it can be scaled to become a repeatable game. For such applications, it can be interesting to study how long effects remain. Insights into this can be used to create more options for repeatability, or give regular updates to make old users return to the application.

With more long term data, aspects such as the motivation to continue with using the system, replayability, and long term effects after using the system could be explored. With larger study populations, it would be possible to also see if effects or experiences are correlated with personal characteristics. This would allow future research to be more personalized, for example by using game mechanics that are effective/appreciated given the user's characteristics.

Another interesting potential direction for future research in the specific domain of empowering vulnerable target groups is to look into vulnerabilities and target groups that are underrepresented in current research. As mentioned in Chapter 2, most of the targeted vulnerabilities are health-related vulnerabilities. Less research has been performed in the direction of for example social or well-being related vulnerability risks. Such risks might need another approach. For example, learning how to deal with feelings of loneliness results in a behaviour that is hard to quantify and therefore harder to directly reward. Whereas behaviours such as healthy moving can be measured by active minutes or numbers of steps which can be linked to rewards. Another problem can be that the topic is more sensitive, and the design requires a very considerate approach. Although working on other topics can be harder, it can have a lot of societal impacts and can thus be worthy to put effort into.

Aside from exploring new vulnerabilities and target groups, it is also interesting to study existing work in more depth. For example, Part III discussed the (as far as I know) first serious game to train verbal resilience using multiple choice and voice analysis. In future research, other games to address the same problem could be tested and compared, to improve the existing serious game. This also holds for the work in the other parts. Combining different researches in the same research direction will lead to more generalized insights, for example into which platforms and game mechanics are suitable for certain groups and vulnerabilities.

It is also interesting to further study the AI components that are described in Chapter 3, 7, and 10. With different cycles of implementing, testing and improving the algorithms, the potential of the components can be further explored. By using the algorithms for personalized meal recommendations (Chapter 7), changes could be made to the diet tracking system. I would be interested to see how the older adult target group would use such recommendations and whether the user-friendliness would benefit from it. It would also be interesting to compare the outcomes of the currently used AI components with other, similar, components. For example, different types of topic algorithms could be compared with the algorithm described in Chapter 10, or other recommendation algorithms that can be used to suggest meals in comparison to the work in Chapter 7.

The data from the evaluation study of the self-compassion training (Chapter 11) can be used to improve the performance of the topic detection algorithm (Chapter 10). An important observation from the preliminary analysis of this larger data set from the evalution study is that there is an imbalance in the gold labels (labels assigned by researchers as ground truth) in this data set. Mostly, this is related to the label ‘no topic’, used for texts that do not discuss one of the predefined topics or discuss multiple topics. In the data set used to build the algorithm, a quatre of the texts was assigned the gold label ‘no topic’. However, the preliminary analysis found that around half of the texts should be assigned the gold label ‘no topic’, mostly because texts discuss multiple topics, but sometimes also because texts are about something that did not fit any of the labels (such as finances). This imbalance in the data set can affect the accuracy and the way that this accuracy should be interpreted in future work: even if the algorithm only predicts ‘no topic’ it will be right in 50% of the cases, thus looking at the accuracies for the topics labels, excluding ‘no topic’ would be a better way to evaluate the algorithm. Moreover, future work should consider how the number of ‘no topic’ gold labels could be reduced since notes with this label cannot be used to personalize the experience in the training. An example of a change to explore is allowing that multiple labels are assigned to one note and only using ‘no topic’ if none of the topics applies. It should also be explored if there are topic labels that can be added or changed based on the larger data set.

Technological improvements can have an impact on the possibilities that serious games and gamified applications have. This can apply to the hardware, which devices and sensors are available, as well as to the used software, such as new possibilities for face and motion capture to make more believable avatars. As a result, it is interesting to keep up with technological innovation to see what new opportunities this gives to existing work, or to work that has not been done as there was not yet a proper way of addressing it. Augmented Reality (AR) is such a technological improvement that can advance the field of serious gaming and gamification. With AR technology the digital world can be integrated into the real world, which has many possible applications. For example, in the future, AR technology could be used for the diet tracking system that I described in Part IV. If users could take AR technology with them during grocery shopping, the application could help them make choices following their diet plan.

## 12.5 Conclusion

To conclude this dissertation, I discuss the main research question:

> *How can serious games and gamification be used to empower vulnerable target groups?*

This question is addressed by the different contributions of this dissertation. First of all, this dissertation contributed to the understanding of the domain. I have identified three different empowerment methods that can be used for vulnerable target groups: training skills, supporting behaviour, and transferring knowledge. The serious game or gamified application has a different role in each of these empowerment methods. Different game mechanics can be used to achieve this. Moreover, requirements for the design are based on the characteristics and wishes of the target group as well as domain-specific requirements. One of the gaps in existing research that I have identified is a lack of a standardized way of describing interventions. The taxonomy in Chapter 2 is an attempt to provide a more standardized vocabulary. I have shown that this taxonomy is sufficient by applying its terms to describe all other parts of this dissertation.

The different parts of this dissertation have shown how development and evaluation are intertwined steps when working on serious games or gamified applications. With this work, I have contributed to research on the development of such applications. I have shown how focus groups, domain experts and domain knowledge can be incorporated into a design phase of an application, which is important to ensure that the application matches with the requirements and wishes of the users. Moreover, it is important to consider whether the effects of game elements do not hinder the desired behavioural outcomes. For example, points and rewards do sometimes have to be limited, if excessive earning is considered not beneficial or harmful. In Parts IV and V, this was important, as the behaviour needed to be optimized instead of maximized: consuming the right amount of protein instead of as much protein as possible and practising with self-compassion regularly instead of as much as possible. In Part IV this was solved by only rewarding the right consumption (with green colouring and a mini-game) and in Part V by giving bonus points after a set number of interactions and limiting the rewards per day. Such limits also help to avoid users getting hooked on using an application, which is often not beneficial for achieving the application's goal.

Serious games and gamified applications can be improved by complementing game mechanics with AI techniques. This offers additional possibilities for features, personalisation and adaptation of content. This dissertation has contributed to showing how different AI techniques can be used for different purposes: e.g. analysing the users' tone of voice, recommending meals, recognizing topics. The added features by the tone of voice analysis and the recognition of topics were both seen as good additions by participants, although their performance was still scarce and needed further improvements. The recommendation of meals could not be evaluated with participants, but results on the performance showed that it lowered the number of clicks needed based on historical data from the diet tracking application. This dissertation is also an example of how knowledge representation is needed to create serious games and applications to empower vulnerable people. For any topic, domain knowledge must be correctly represented in the application. This was particularly important for the work of Part IV where the domain knowledge about the diet needed to be precisely represented in the application. In this process, it became apparent that such domain knowledge is sometimes open to human interpretation (for example in rounding) which makes knowledge representation a difficult task.

Finally, this dissertation has contributed to three different domains: safety for older adults, healthy nutrition for older adults, and mental well-being for young adults. The results from the studies described in this dissertation provide information that can be used for future applications in these domains. In addition, the results found in Part III and Part IV contribute to research towards technology or gamification adoption of older adults. For example, the results described in Chapter 6 show that older adults are very adherent users of a diet tracking system within a diet trial.

Serious games and gamification are tools that can be used to empower vulnerable people in many different ways. An important advantage of such tools over human-controlled interventions is that they can be easily scaled and lower barriers for participants to join. However, digital interventions such as described in this dissertation do not have to eliminate in-person activities. As shown in the work of Part III on doorstep scams prevention, participants also see the added value of combining digital tools (in this case a serious game) with an in-person aspect (such as an information meeting). In the work on the self-compassion training program (Part V), I have discussed that a possible solution to reduce the drop-out rate is including an in-person component, such as professional supervision. In the work described in Part IV on diet tracking for older adults, such an in-person component was present, as participants were part of a diet trial and were also in contact with a dietitian. High adherence was found in this study, however, it could not be studied if this was due to the in-person component as there was no group using the tablet without the in-person contact. Although serious games and gamified applications are often seen as private activities, adding a social component can add value to the experience. This does not necessarily mean offering social activities with other users, but could also be using an application in a group session or with some form of professional supervision.

This dissertation contributes to understanding the domain of serious games and gamification. I have shown the diversity of the domain, but also emphasized what is shared. One of the lessons learned is that to design meaningful gamification, game mechanics should carefully be balanced. Matching the game mechanics with the domain knowledge and requirements of users and other stakeholders, improves the user friendliness, supports the effects, and minimizes misuse. Another lesson learned is that adding AI features allows personalization of the experience or feedback. Advances in the domain will hopefully empower a broad range of vulnerable target groups.



## Bibliography

[1] Al-Razgan, M.S., Al-Khalifa, H.S., Al-Shahrani, M.D., AlAjmi, H.H.: Touch-based mobile phone interface guidelines and design recommendations for elderly people: A survey of the literature. In: Int. Conf. Neural Inform. Processing. Springer (2012)

[2] Hellman, R.: Universal design and mobile devices. In: International Conference on Universal Access in Human-Computer Interaction. pp. 147–156. Springer (2007)

[3] Jones, S., Hara, S., Augusto, J.C.: efriend: an ethical framework for intelligent environments development. Ethics and information technology **17**(1), 11–25 (2015)

[4] Koivisto, J., Hamari, J.: Demographic differences in perceived benefits from gamification. Computers in Human Behavior **35**, 179–188 (2014)

[5] Neff, K.D., Germer, C.K.: A pilot study and randomized controlled trial of the mindful self-compassion program. Journal of clinical psychology **69**(1), 28–44 (2013). https://doi.org/10.1037/0022-3514.92.5.887

[6] Reinders, I., Visser, M., Jyväkorpi, S.K., Niskanen, R.T., Bosmans, J.E., Jornada Ben, Â., Brouwer, I.A., Kuijper, L.D., Olthof, M.R., Pitkälä, K.H., et al.: The cost effectiveness of personalized dietary advice to increase protein intake in older adults with lower habitual protein intake: a randomized controlled trial. European journal of nutrition pp. 1–16 (2021)

[7] Schmidt-Kraepelin, M., Thiebes, S., Stepanovic, S., Mettler, T., Sunyaev, A.: Gamification in health behavior change support systems-a synthesis of unintended side effects. In: Proceedings of the 14th International Conference on Wirtschaftsinformatik. pp. 1032–1046 (2019)

[8] Thomas, R.J., Masthoff, J., Oren, N.: Adapting healthy eating messages to personality. In: Int. Conf. on Persuasive Technology. pp. 119–132. Springer (2017)

[9] Van Genugten, L., Dusseldorp, E., Webb, T.L., Van Empelen, P.: Which combinations of techniques and modes of delivery in internet-based interventions effectively change health behavior? a meta-analysis. Journal of medical Internet research **18**(6), e155 (2016)

[10] Webb, T., Joseph, J., Yardley, L., Michie, S.: Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. Journal of medical Internet research **12**(1) (2010)

# Addendum

# List of SIKS Dissertations

## 2016

**Syed Saiden Abbas (RUN)**, Recognition of Shapes by Humans and Machines

**Michiel Christiaan Meulendijk (UU)**, Optimizing medication reviews through decision support: prescribing a better pill to swallow

**Maya Sappelli (RUN)**, Knowledge Work in Context: User Centered Knowledge Worker Support

**Laurens Rietveld (VU)**, Publishing and Consuming Linked Data

**Evgeny Sherkhonov (UVA)**, Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers

**Michel Wilson (TUD)**, Robust scheduling in an uncertain environment

**Jeroen de Man (VU)**, Measuring and modeling negative emotions for virtual training

**Matje van de Camp (TiU)**, A Link to the Past: Constructing Historical Social Networks from Unstructured Data

**Archana Nottamkandath (VU)**, Trusting Crowdsourced Information on Cultural Artefacts

**George Karafotias (VUA)**, Parameter Control for Evolutionary Algorithms

**Anne Schuth (UVA)**, Search Engines that Learn from Their Users

**Max Knobbout (UU)**, Logics for Modelling and Verifying Normative Multi-Agent Systems

**Nana Baah Gyan (VU)**, The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach

**Ravi Khadka (UU)**, Revisiting Legacy Software System Modernization

**Steffen Michels (RUN)**, Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments

**Guangliang Li (UVA)**, Socially Intelligent Autonomous Agents that Learn from Human Reward

**Berend Weel (VU)**, Towards Embodied Evolution of Robot Organisms

**Albert Meroño Peñuela (VU)**, Refining Statistical Data on the Web

**Julia Efremova (Tu/e)**, Mining Social Structures from Genealogical Data

**Daan Odijk (UVA)**, Context & Semantics in News & Web Search

**Alejandro Moreno Célleri (UT)**, From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground

**Grace Lewis (VU)**, Software Architecture Strategies for Cyber-Foraging Systems

**Fei Cai (UVA)**, Query Auto Completion in Information Retrieval

**Brend Wanders (UT)**, Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach

**Julia Kiseleva (TU/e)**, Using Contextual Information to Understand Searching and Browsing Behavior

**Dilhan Thilakarathne (VU)**, In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains

**Wen Li (TUD)**, Understanding Geo-spatial Information on Social Media

**Mingxin Zhang (TUD)**, Large-scale Agent-based Social Simulation - A study on epidemic prediction and control

**Nicolas Höning (TUD)**, Peak reduction in decentralised electricity systems - Markets and prices for flexible planning

**Ruud Mattheij (UvT)**, The Eyes Have It

**Mohammad Khelghati (UT)**, Deep web content monitoring

**Eelco Vriezekolk (UT)**, Assessing Telecommunication Service Availability Risks for Crisis Organisations

**Peter Bloem (UVA)**, Single Sample Statistics, exercises in learning from just one example

**Dennis Schunselaar (TUE)**, Configurable Process Trees: Elicitation, Analysis, and Enactment

**Zhaochun Ren (UVA)**, Monitoring Social Media: Summarization, Classification and Recommendation

**Daphne Karreman (UT)**, Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies

**Giovanni Sileno (UvA)**, Aligning Law and Action - a conceptual and computational inquiry

**Andrea Minuto (UT)**, Materials that Matter - Smart Materials meet Art & Interaction Design

**Merijn Bruijnes (UT)**, Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect

**Christian Detweiler (TUD)**, Accounting for Values in Design

**Thomas King (TUD)**, Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance

**Spyros Martzoukos (UVA)**, Combinatorial and Compositional Aspects of Bilingual Aligned Corpora

**Saskia Koldijk (RUN)**, Context-Aware Support for Stress Self-Management: From Theory to Practice

**Thibault Sellam (UVA)**, Automatic Assistants for Database Exploration

**Bram van de Laar (UT)**, Experiencing Brain-Computer Interface Control

**Jorge Gallego Perez (UT)**, Robots to Make you Happy

**Christina Weber (UL)**, Real-time foresight - Preparedness for dynamic innovation networks

**Tanja Buttler (TUD)**, Collecting Lessons Learned

**Gleb Polevoy (TUD)**, Participation and Interaction in Projects. A Game-Theoretic Analysis

**Yan Wang (UVT)**, The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains

## 2017

**Jan-Jaap Oerlemans (UL)**, Investigating Cybercrime

**Sjoerd Timmer (UU)**, Designing and Understanding Forensic Bayesian Networks using Argumentation

**Daniël Harold Telgen (UU)**, Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines

**Mrunal Gawade (CWI)**, Multi-core Parallelism in a Column-store

**Mahdieh Shadi (UVA)**, Collaboration Behavior

**Damir Vandic (EUR)**, Intelligent Information Systems for Web Product Search

**Roel Bertens (UU)**, Insight in Information: from Abstract to Anomaly

**Rob Konijn (VU)** , Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery

**Dong Nguyen (UT)**, Text as Social and Cultural Data: A Computational Perspective on Variation in Text

**Robby van Delden (UT)**, (Steering) Interactive Play Behavior

**Florian Kunneman (RUN)**, Modelling patterns of time and emotion in Twitter #anticipointment

**Sander Leemans (TUE)**, Robust Process Mining with Guarantees

**Gijs Huisman (UT)**, Social Touch Technology - Extending the reach of social touch through haptic technology

**Shoshannah Tekofsky (UvT)**, You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior
**Peter Berck (RUN)**, Memory-Based Text Correction
**Aleksandr Chuklin (UVA)**, Understanding and Modeling Users of Modern Search Engines
**Daniel Dimov (UL)**, Crowdsourced Online Dispute Resolution
**Ridho Reinanda (UVA)**, Entity Associations for Search
**Jeroen Vuurens (UT)**, Proximity of Terms, Texts and Semantic Vectors in Information Retrieval
**Mohammadbashir Sedighi (TUD)**, Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility
**Jeroen Linssen (UT)**, Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds)
**Sara Magliacane (VU)**, Logics for causal inference under uncertainty
**David Graus (UVA)**, Entities of Interest — Discovery in Digital Traces
**Chang Wang (TUD)**, Use of Affordances for Efficient Robot Learning
**Veruska Zamborlini (VU**), Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search
**Merel Jung (UT)**, Socially intelligent robots that understand and respond to human touch
**Michiel Joosse (UT)**, Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behaviors
**John Klein (VU)**, Architecture Practices for Complex Contexts
**Adel Alhuraibi (UvT)**, From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT"
**Wilma Latuny (UvT)**, The Power of Facial Expressions
**Ben Ruijl (UL)**, Advances in computational methods for QFT calculations
**Thaer Samar (RUN)**, Access to and Retrievability of Content in Web Archives
**Brigit van Loggem (OU)**, Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity
**Maren Scheffel (OU)**, The Evaluation Framework for Learning Analytics
**Martine de Vos (VU)**, Interpreting natural science spreadsheets
**Yuanhao Guo (UL)**, Shape Analysis for Phenotype Characterisation from High-throughput Imaging
**Alejandro Montes Garcia (TUE)**, WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy
**Alex Kayal (TUD)**, Normative Social Applications
**Sara Ahmadi (RUN)**, Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR
**Altaf Hussain Abro (VUA)**, Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems
**Adnan Manzoor (VUA)**, Minding a Healthy Lifestyle: An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle
**Elena Sokolova (RUN)**, Causal discovery from mixed and missing data with applications on ADHD datasets
**Maaike de Boer (RUN)**, Semantic Mapping in Video Retrieval
**Garm Lucassen (UU)**, Understanding User Stories - Computational Linguistics in Agile Requirements Engineering
**Bas Testerink (UU)**, Decentralized Runtime Norm Enforcement
**Jan Schneider (OU)**, Sensor-based Learning Support
**Jie Yang (TUD)**, Crowd Knowledge Creation Acceleration
**Angel Suarez (OU)**, Collaborative inquiry-based learning

## 2018

**Han van der Aa (VUA)**, Comparing and Aligning Process Representations
**Felix Mannhardt (TUE)**, Multi-perspective Process Mining
**Steven Bosems (UT)**, Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction
**Jordan Janeiro (TUD)**, Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks
**Hugo Huurdeman (UVA)**, Supporting the Complex Dynamics of the Information Seeking Process
**Dan Ionita (UT)**, Model-Driven Information Security Risk Assessment of Socio-Technical Systems
**Jieting Luo (UU)**, A formal account of opportunism in multi-agent systems
**Rick Smetsers (RUN)**, Advances in Model Learning for Software Systems
**Xu Xie (TUD)**, Data Assimilation in Discrete Event Simulations
**Julienka Mollee (VUA)**, Moving forward: supporting physical activity behavior change through intelligent technology
**Mahdi Sargolzaei (UVA)**, Enabling Framework for Service-oriented Collaborative Networks
**Xixi Lu (TUE)**, Using behavioral context in process mining
**Seyed Amin Tabatabaei (VUA)**, Computing a Sustainable Future
**Bart Joosten (UVT)**, Detecting Social Signals with Spatiotemporal Gabor Filters
**Naser Davarzani (UM)**, Biomarker discovery in heart failure
**Jaebok Kim (UT)**, Automatic recognition of engagement and emotion in a group of children
**Jianpeng Zhang (TUE)**, On Graph Sample Clustering
**Henriette Nakad (UL)**, De Notaris en Private Rechtspraak
**Minh Duc Pham (VUA)**, Emergent relational schemas for RDF
**Manxia Liu (RUN)**, Time and Bayesian Networks
**Aad Slootmaker (OUN)**, EMERGO: a generic platform for authoring and playing scenario-based serious games
**Eric Fernandes de Mello Araujo (VUA)**, Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks
**Kim Schouten (EUR)**, Semantics-driven Aspect-Based Sentiment Analysis
**Jered Vroon (UT)**, Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots
**Riste Gligorov (VUA)**, Serious Games in Audio-Visual Collections
**Roelof Anne Jelle de Vries (UT)**,Theory-Based and Tailor-Made: Motivational Messages for Behavior Change Technology
**Maikel Leemans (TUE)**, Hierarchical Process Mining for Scalable Software Analysis
**Christian Willemse (UT)**, Social Touch Technologies: How they feel and how they make you feel
**Yu Gu (UVT)**, Emotion Recognition from Mandarin Speech
**Wouter Beek**, The "K" in "semantic web" stands for "knowledge": scaling semantics to the web

## 2019

**Rob van Eijk (UL)**, Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification
**Emmanuelle Beauxis Aussalet (CWI, UU)**, Statistics and Visualizations for Assessing Class Size Uncertainty
**Eduardo Gonzalez Lopez de Murillas (TUE)**, Process Mining on Databases: Extracting Event Data from Real Life Data Sources
**Ridho Rahmadi (RUN)**, Finding stable causal structures from clinical data

**Sebastiaan van Zelst (TUE)**, Process Mining with Streaming Data

**Chris Dijkshoorn (VU)**, Nichesourcing for Improving Access to Linked Cultural Heritage Datasets

**Soude Fazeli (TUD)**, Recommender Systems in Social Learning Platforms

**Frits de Nijs (TUD)**, Resource-constrained Multi-agent Markov Decision Processes

**Fahimeh Alizadeh Moghaddam (UVA)**, Self-adaptation for energy efficiency in software systems

**Qing Chuan Ye (EUR)**, Multi-objective Optimization Methods for Allocation and Prediction

**Yue Zhao (TUD)**, Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs

**Jacqueline Heinerman (VU)**, Better Together

**Guanliang Chen (TUD)**, MOOC Analytics: Learner Modeling and Content Generation

**Daniel Davis (TUD)**, Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses

**Erwin Walraven (TUD)**, Planning under Uncertainty in Constrained and Partially Observable Environments

**Guangming Li (TUE)**, Process Mining based on Object-Centric Behavioral Constraint (OCBC) Models

**Ali Hurriyetoglu (RUN)**,Extracting actionable information from microtexts

**Gerard Wagenaar (UU)**, Artefacts in Agile Team Communication

**Vincent Koeman (TUD)**, Tools for Developing Cognitive Agents

**Chide Groenouwe (UU)**, Fostering technically augmented human collective intelligence

**Cong Liu (TUE)**, Software Data Analytics: Architectural Model Discovery and Design Pattern Detection

**Martin van den Berg (VU)**,Improving IT Decisions with Enterprise Architecture

**Qin Liu (TUD)**, Intelligent Control Systems: Learning, Interpreting, Verification

**Anca Dumitrache (VU)**, Truth in Disagreement - Crowdsourcing Labeled Data for Natural Language Processing

**Emiel van Miltenburg (VU)**, Pragmatic factors in (automatic) image description

**Prince Singh (UT)**, An Integration Platform for Synchromodal Transport

**Alessandra Antonaci (OUN)**, The Gamification Design Process applied to (Massive) Open Online Courses

**Esther Kuindersma (UL)**, Cleared for take-off: Game-based learning to prepare airline pilots for critical situations

**Daniel Formolo (VU)**, Using virtual agents for simulation and training of social skills in safety-critical circumstances

**Vahid Yazdanpanah (UT)**, Multiagent Industrial Symbiosis Systems

**Milan Jelisavcic (VU)**, Alive and Kicking: Baby Steps in Robotics

**Chiara Sironi (UM)**, Monte-Carlo Tree Search for Artificial General Intelligence in Games

**Anil Yaman (TUE)**, Evolution of Biologically Inspired Learning in Artificial Neural Networks

**Negar Ahmadi (TUE)**, EEG Microstate and Functional Brain Network Features for Classification of Epilepsy and PNES

**Lisa Facey-Shaw (OUN)**, Gamification with digital badges in learning programming

**Kevin Ackermans (OUN)**, Designing Video-Enhanced Rubrics to Master Complex Skills

**Jian Fang (TUD)**, Database Acceleration on FPGAs

**Akos Kadar (OUN)**, Learning visually grounded and multilingual representations

## 2020

**Armon Toubman (UL)**, Calculated Moves: Generating Air Combat Behaviour

**Marcos de Paula Bueno (UL)**, Unraveling Temporal Processes using Probabilistic Graphical Models

**Mostafa Deghani (UvA)**, Learning with Imperfect Supervision for Language Understanding
**Maarten van Gompel (RUN)**, Context as Linguistic Bridges
**Yulong Pei (TUE)**, On local and global structure mining
**Preethu Rose Anish (UT)**, Stimulation Architectural Thinking during Requirements Elicitation - An Approach and Tool Support
**Wim van der Vegt (OUN)**, Towards a software architecture for reusable game components
**Ali Mirsoleimani (UL)**,Structured Parallel Programming for Monte Carlo Tree Search
**Myriam Traub (UU)**, Measuring Tool Bias and Improving Data Quality for Digital Humanities Research
**Alifah Syamsiyah (TUE)**, In-database Preprocessing for Process Mining
**Sepideh Mesbah (TUD)**, Semantic-Enhanced Training Data AugmentationMethods for Long-Tail Entity Recognition Models
**Ward van Breda (VU)**, Predictive Modeling in E-Mental Health: Exploring Applicability in Personalised Depression Treatment
**Marco Virgolin (CWI)**, Design and Application of Gene-pool Optimal Mixing Evolutionary Algorithms for Genetic Programming
**Mark Raasveldt (CWI/UL)**, Integrating Analytics with Relational Databases
**Konstantinos Georgiadis (OUN)**, Smart CAT: Machine Learning for Configurable Assessments in Serious Games
**Ilona Wilmont (RUN)**, Cognitive Aspects of Conceptual Modelling
**Daniele Di Mitri (OUN)**, The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences
**Georgios Methenitis (TUD)**, Agent Interactions & Mechanisms in Markets with Uncertainties: Electricity Markets in Renewable Energy Systems
**Guido van Capelleveen (UT)**, Industrial Symbiosis Recommender Systems
**Albert Hankel (VU)**, Embedding Green ICT Maturity in Organisations
**Karine da Silva Miras de Araujo (VU)**, Where is the robot?: Life as it could be
**Maryam Masoud Khamis (RUN)**, Understanding complex systems implementation through a modeling approach: the case of e-government in Zanzibar
**Rianne Conijn (UT)**, The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging
**Lenin da Nobrega Medeiros (VUA/RUN)**, How are you feeling, human? Towards emotionally supportive chatbots
**Xin Du (TUE)**, The Uncertainty in Exceptional Model Mining
**Krzysztof Leszek Sadowski (UU)**, GAMBIT: Genetic Algorithm for Model-Based mixed-Integer opTimization
**Ekaterina Muravyeva (TUD)**, Personal data and informed consent in an educational context
**Bibeg Limbu (TUD)**, Multimodal interaction for deliberate practice: Training complex skills with augmented reality
**Ioan Gabriel Bucur (RUN)**, Being Bayesian about Causal Inference
**Bob Zadok Blok (UL)**, Creatief, Creatieve, Creatiefst
**Gongjin Lan (VU)**, Learning better – From Baby to Better
**Jason Rhuggenaath (TUE)**, Revenue management in online markets: pricing and online advertising
**Rick Gilsing (TUE)**, Supporting service-dominant business model evaluation in the context of business model innovation
**Anna Bon (MU)**, Intervention or Collaboration? Redesigning Information and Communication Technologies for Development
**Siamak Farshidi (UU)**, Multi-Criteria Decision-Making in Software Production

## 2021

**Francisco Xavier Dos Santos Fonseca (TUD)**,Location-based Games for Social Interaction in Public Space

**Rijk Mercuur (TUD)**, Simulating Human Routines:Integrating Social Practice Theory in Agent-Based Models

**Seyyed Hadi Hashemi (UVA)**, Modeling Users Interacting with Smart Devices

**Ioana Jivet (OU)**, The Dashboard That Loved Me: Designing adaptive learning analytics for self-regulated learning

**Davide Dell'Anna (UU)**, Data-Driven Supervision of Autonomous Systems

**Daniel Davison (UT)**, "Hey robot, what do you think?" How children learn with a social robot

**Armel Lefebvre (UU)**, Research data management for open science

**Nardie Fanchamps (OU)**, The Influence of Sense-Reason-Act Programming on Computational Thinking

**Cristina Zaga (UT)**, The Design of Robothings. Non-Anthropomorphic and Non-Verbal Robots to Promote Children's Collaboration Through Play

**Quinten Meertens (UvA)**, Misclassification Bias in Statistical Learning

**Anne van Rossum (UL)**, Nonparametric Bayesian Methods in Robotic Vision

**Lei Pi (UL)**, External Knowledge Absorption in Chinese SMEs

**Bob R. Schadenberg (UT)**, Robots for Autistic Children: Understanding and Facilitating Predictability for Engagement in Learning

**Negin Samaeemofrad (UL)**, Business Incubators: The Impact of Their Support

**Onat Ege Adali (TU/e)**, Transformation of Value Propositions into Resource Re-Configurations through the Business Services Paradigm

**Esam A. H. Ghaleb (UM)**, BIMODAL EMOTION RECOGNITION FROM AUDIO-VISUAL CUES

**Dario Dotti (UM)**, Human Behavior Understanding from motion and bodily cues using deep neural networks

**Remi Wieten (UU)**, Bridging the Gap Between Informal Sense-Making Tools and Formal Systems - Facilitating the Construction of Bayesian Networks and Argumentation Frameworks

**Roberto Verdecchia (VU)**, Architectural Technical Debt: Identification and Management

**Masoud Mansoury (TU/e)**, Understanding and Mitigating Multi-Sided Exposure Bias in Recommender Systems

**Pedro Thiago Timbó Holanda (CWI)**, Progressive Indexes

**Sihang Qiu (TUD)**, Conversational Crowdsourcing

**Hugo Manuel Proença (LIACS)**, Robust rules for prediction and description

**Kaijie Zhu (TUE)**, On Efficient Temporal Subgraph Query Processing

**Eoin Martino Grua (VUA)**, The Future of E-Health is Mobile: Combining AI and Self-Adaptation to Create Adaptive E-Health Mobile Applications

**Benno Kruit (CWI & VUA)**, Reading the Grid: Extending Knowledge Bases from Human-readable Tables

**Jelte van Waterschoot (UT)**, Personalized and Personal Conversations: Designing Agents Who Want to Connect With You

**Christoph Selig (UL)**, Understanding the Heterogeneity of Corporate Entrepreneurship Programs

## 2022

**Judith van Stegeren (UT)**, Flavor text generation for role-playing video games

**Paulo da Costa (TU/e)**, Data-driven Prognostics and Logistics Optimisation: A Deep Learning Journey

**Ali el Hassouni (VUA)**, A Model A Day Keeps The Doctor Away: Reinforcement Learning For Personalized Healthcare
**Ünal Aksu (UU)**, A Cross-Organizational Process Mining Framework
**Shiwei Liu (TU/e)**, Sparse Neural Network Training with In-Time Over-Parameterization
**Reza Refaei Afshar (TU/e)**, Machine Learning for Ad Publishers in Real Time Bidding
**Sambit Praharaj (OU)**, Measuring the Unmeasurable? Towards Automatic Co-located Collaboration Analytics
**Maikel L. van Eck (TU/e)**, Process Mining for Smart Product Design
**Oana Andreea Inel (VUA)**, Understanding Events: A Diversity-driven Human-Machine Approach
**Felipe Moraes Gomes (TUD)**, Examining the Effectiveness of Collaborative Search Engines
**Mirjam de Haas (UT)**, Staying engaged in child-robot interaction, a quantitative approach to studying preschoolers' engagement with robots and tasks during second-language tutoring
**Guanyi Chen (UU)**, Computational Generation of Chinese Noun Phrases
**Xander Wilcke (VUA)**, Machine Learning on Multimodal Knowledge Graphs: Opportunities, Challenges, and Methods for Learning on Real-World Heterogeneous and Spatially-Oriented Knowledge
**Michiel Overeem (UU)**, Evolution of Low-Code Platforms
**Jelmer Jan Koorn (UU)**, Work in Process: Unearthing Meaning using Process Mining