Improving the Grow It! M-Health Tool: Examining the Role of Real-Time Personalized Feedback

Meyser, Louise

Universiteit Leiden

Psychologie
Faculteit der Sociale Wetenschappen

Improving the Grow It! M-Health Tool: Examining the Role of Real-Time Personalized Feedback

Louise Meyser

Master Thesis Clinical Neuropsychology

Faculty of Behavioral and Social Sciences – Leiden University

June 2023

Student number: s2061775

External Supervisor: Evelien Dietvorst & Anne Snijders, Erasmus MC Sophia

First Examiner: Nicholas Root, Health, Medical and Neuropsychology Unit; Leiden University

Abstract

Background: Studying the effects of mobile health (m-health) apps is crucial for providing accessible mental healthcare to people at high risk for psychopathology, such as adolescents during the COVID-19 pandemic. Therefore, investigating elements such as real-time personalized feedback are essential to understand what underpins the effectiveness of m-health. The Grow It! app is a multiplayer serious gaming m-health app for adolescents aged 12 to 25 aimed at identification of emotional problems and improving well-being. This study aimed to improve suboptimal app activity (i.e., amount of experience sampling method questionnaires completed) by incorporating feedback in the form of an emotion overview chart into the app. Furthermore, this study wanted to investigate changes in affective and cognitive well-being and assess how users rate the emotion overview. **Method:** Adolescents ($N = 143$) played Grow It! for three weeks during the COVID-19 pandemic. Participants filled in a questionnaire before and after playing the app, to measure differences in affective (measured on a 7-point Likert scale) and cognitive (measured on a 10-point Likert scale) well-being. The emotion overview was evaluated with four questions. **Results:** After conducting two paired samples t-tests, I found that affective well-being significantly increased by 0.29 points, $t(142) = 3.30$, $p < .001$ (one-tailed), $d = .28$. Forty percent (40%) of individuals experienced increases. Cognitive well-being significantly increased with 0.44 points, $t(142) = 3.17$, $p < .001$ (one-tailed), $d = .27$. Forty-nine percent (49%) of individuals experienced increases. After conducting a one-way ANOVA, I found that app activity was significantly higher for users playing Grow It! with the emotion overview included, $F(2,1335) = 53.13$, $p < .001$. User evaluations were overall positive with the emotion overview being a welcomed addition. **Conclusion:** The results of this study demonstrate successful replication of previous studies and support gamification theories on the usefulness of real-time personalized feedback. The addition of the emotion overview seems valuable for increasing app activity and positive app evaluations. Further research with a control group is recommended to be able to make any substantial claims on the effectiveness of Grow It! and what added value the emotion overview provides.

Keywords: Grow It!, Experience Sampling Method, Personalized Feedback, Emotion Overview, Compliance, Engagement, App Activity, M-health Application, Serious Gaming Application, Cognitive Well-being, Affective Well-being, User Evaluation.

Layman's Abstract

Background: This study investigates the effects of the mobile health (m-health) app called Grow It! on the well-being of young adults during the COVID-19 pandemic. Researchers designed Grow It! to help identify emotional problems and improve overall well-being in adolescents. The goal of this study was to understand if incorporating feedback (in the form of an emotion overview chart) that is timely and tailored to the individual has any effect on the effectiveness of the app. In particular, I aimed to investigate whether adolescents would improve in their mental and emotional well-being, whether including feedback would increase app usage and how users value its addition. **Methods:** Over a period of three weeks 143 adolescents played Grow It!, before and after which they filled in questionnaires concerning their well-being and evaluation of the app. App users evaluated the emotion overview with four questions. **Results:** The results showed that adolescents improved in their mental and emotional well-being after using Grow It! with the emotion overview included. Almost half of the adolescents had significant improvements. Adolescents rated Grow It! and highly and the emotion overview was a welcomed addition. The app usage was higher in comparison to other studies on Grow It!. **Conclusion:** These findings suggest that including real-time personalized feedback, such as the emotion overview chart, may improve the app experience and increase user engagement with Grow It!. This supports past research that state that feedback helps increase user interaction. However, future research requires a control group to determine whether changes in well-being are not due to any other factors other than the app, to truly be able to make any large claims about the effectiveness of Grow It! and what role the emotion overview plays.

Contents

Improving the Grow It! M-Health Tool: Examining the Role of Real-Time Personalized Feedback

Mobile health (m-health) apps are increasingly popular and accessible tools that can aid the early identification, prevention, and treatment of mental health problems (Clarke et al., 2015). Free and colleagues (2010) broadly define m-health as "the use of mobile computing and communication technologies in health care and public health". Benefits of m-health apps include that they can help ease the burden on healthcare systems (Bakker et al., 2016) and adolescents perceive them as less stigmatizing than traditional therapy (Bergin et al., 2020). Amidst the recent COVID-19 pandemic, the importance of m-health apps has heightened. The pandemic has led to the temporary closing down of schools and limited sports opportunities, reducing opportunities for in-person social interaction and connection. The consequences for young adults are great as adolescence is a crucial developmental period for gaining autonomy from parents while developing peer relationships becomes the focal point (Gorrese & Ruggieri, 2012; March-Llanes et al., 2017). Health risks associated with social isolation and loneliness are enormous (Friedler et al., 2015), putting many adolescents living in times of a pandemic at an increased risk for developing psychopathology. Therefore, adolescents could potentially benefit from something as accessible as m-health. Nevertheless, elements applied in m-health that underpin the effectiveness of such interventions remain unclear and require a better understanding to give at-risk adolescents the help they need.

Feedback may be a crucial component of effective m-health, as supported by compelling arguments from therapeutic and gaming perspectives. From a therapeutic perspective, feedback plays a vital role in increasing self-reflection and self-monitoring, encouraging help-seeking and motivating patients towards behavioral activation (Bakker et al., 2016). By self-reporting thoughts, feelings and behaviors patients can gain a deeper understanding of their mental health. Feedback provides an extra layer of self-awareness, serving as a catalyst for action. In addition, feedback prevents recall bias allowing for an accurate representation of experiences, encouraging individuals to confront their true behaviors and seek appropriate help. Feedback can take various forms. Therapists commonly provide verbal feedback regarding patients' treatment progress in cognitive behavioral therapy (CBT; Janse et al., 2020). This feedback is particularly effective when tailored to the individuals needs and preferences as individual differences are considered (Bastiaansen et al., 2018). Another effective form of feedback is visual feedback such as charts and graphs. The visual form provides a new perspective and aids cognitive processing as complex information is simplified and certain information becomes more explicit, creating a more complete understanding (Bobek & Tversky, 2016; Kramer et al., 2014). In addition, visual feedback may increase engagement and motivation.

In the realm of gaming, feedback plays a key role in the progression mechanics inherent to video games. The progression mechanic principle highlights that receiving feedback plays a role in improving intrinsic motivation to engage. (Cugelman, 2013; Robson et al., 2015). Take for example a game with increasing levels of difficulty. As players master each level, they feel a sense of accomplishment and are motivated to continue. Feedback on the progression through various levels functions as a reward, reinforcing positive behaviors. Feedback on player's progression also promotes learning. Learning about game mechanics helps players to understand the game and optimize their performance (Burgos et al., 2007; Laamarti et al., 2014). When a player achieves a particular balance between the challenge of the game and their skill level, they enter a state of "flow". Flow is a state of deep enjoyment, complete immersion, and focused engagement (Csikszentmihalyi, 1990), heightened through immediate and meaningful feedback loops (Cowley et al., 2008). These feedback loops help create seamless interactions between the actions of the player and the game's response, enhancing player's connection to a game world. Overall, the aforementioned literature demonstrates the benefit and broad applicability of feedback in multiple contexts. Therapy and gaming objectives come together in serious games; games that prioritizing therapeutic purposes beyond entertainment (Laamarti et al., 2014). Moreover, serious games may align particularly well with the needs of the current generations of digital natives, who have grown up in an era of widespread online gaming. Therefore, serious games provide a relevant way to investigate the potential of feedback in m-health tools for adolescents.

The Grow It! App

The current research study focusses on the Grow It! app. In short, Grow It! is a multiplayer serious gaming m-health app specifically designed for adolescents aged 12 to 25 aimed at early identification of emotional problems and improving well-being through promoting self-insight and self-monitoring of moods and behaviors with experience sampling method (ESM; Myin-Germeys et al., 2018) questionnaires. Furthermore, Grow It! aims to promote adaptive coping underpinned by CBT principles. The app is gamified by means of a point earning system and competition against other (anonymous) players. A recent publication by Dietvorst et al., (2022a) revealed that adolescents showed significant increases in cognitive and affective well-being after playing Grow It! for three to six weeks during the COVID-19 pandemic. The authors corrected their results for confounders; however, any big conclusions cannot be drawn due to a lack of a control group. Furthermore, Dietvorst et al. (2022b) conducted a feasibility study for the development of Grow It!. Findings were mixed. On the one hand users evaluated the app positively, while on the other hand the number of responses to the ESM questionnaires were limited, demonstrating suboptimal app activity. The adolescents in this study completed on average one third of all the ESM questionnaires, at most. The authors highlight the potential benefit of adding a visual feedback component. As a result, we have created an "extended" version of Grow It!. It includes real-time personalized feedback in the form of an emotion chart, where the user gets visual feedback on their emotions in specific situations (see Figure 1). The goal of including this feedback element is to improve the effectiveness of the ESM questionnaires to increase app activity and enhance therapeutic effects. Therapeutic effects may indirectly contribute to more app activity as well (Dietvorst et al., 2022b).

Figure 1

Illustration of the Experience Sampling Method (ESM) Component in Grow It! Extended

Note. The visualization on the left represents a question from the ESM questionnaire. A user must fill in how relaxed they feel in this particular moment, rated on a scale from “not” to “very”. The visualization on the right represents the emotion overview. In this example the user views their mood on Thursday. The bigger the bubble the stronger they experienced the emotion.

Study Objectives

For this thesis three research objectives are preregistered (see Appendix A). Firstly, (RO_1) this study aims to investigate whether affective and cognitive well-being in adolescents improve after playing Grow It! extended (Replication following Dietvorst et al., 2022a). I formulated two hypotheses: H_{1A}: Adolescents will show an increase in their affective well-being after playing Grow It! extended. H_{1B}: Adolescents will show an increase in their cognitive well-being after playing Grow It! extended. These hypotheses are based on the findings by Dietvorst et al. (2022a) who demonstrate significant increases in affective as well as cognitive well-being after playing the Grow It! original app. This objective adds to the current literature on Grow It! by replicating and supporting previously found effects. In addition, ($RO_{1,SUB}$) this study will describe according to Grice's person-centered effect sizes (2020) what proportion of adolescents change (0.2 *SD*) in their affective and cognitive well-being after playing Grow It! extended (replication following Dietvorst et al., 2022a). The addition of this sub-objective is of exploratory nature and is useful to better help identify how many adolescents benefit from the app on an individual level besides effects predicted to be found at the group level.

(RO_2) Furthermore, this study will investigate whether users who play Grow It! extended will complete more ESM questionnaires than users who played Grow It! original (Dietvorst et al., 2022a). I predicted the following: H_2: Users of Grow It! extended will show more app activity (i.e., amount of completed ESM questionnaires) than users of Grow It! original. Gamification principles that state that providing personalized feedback increases engagement (Robson et al., 2015) support this hypothesis. In addition, providing personalized feedback on mood may enhance therapeutic effects (Dietvorst et al., 2022a), which could lead to more app engagement in an indirect manner. By means of the implementation of the personalized feedback this study tries to improve player activity.

(RO_3) Lastly, this study aims to describe how adolescents evaluate the emotion chart in Grow It! extended and whether its addition has an impact on user evaluations. This objective is of an exploratory nature and is useful to better understand how adolescents experience the implementation of the real-time personalized feedback. By investigating the individual opinions of adolescents on the emotion chart, this study can provide an understanding regarding its usability and effectiveness within the app.

This study adds to m-health literature by examining the role of personalized feedback that Bakker et al. (2016) reports as missing in their systematic literature review on recommendations for creating effective mental health smartphone apps. The findings to the previously mentioned objectives have practical and fundamental implications for the digital health field. On a practical level, the implications of these findings could aid the development and improvement of Grow It!, so that eventually the application can be most effective and accessible for adolescents who require additional support in their mental health and well-being. On a fundamental level, findings can contribute to distinguishing which elements of m-health apps are key in their effectiveness and which ones are perhaps less crucial than initially thought. This knowledge can guide future m-health development efforts with evidence-based decisions and facilitate the continuous advancement of the digital health landscape.

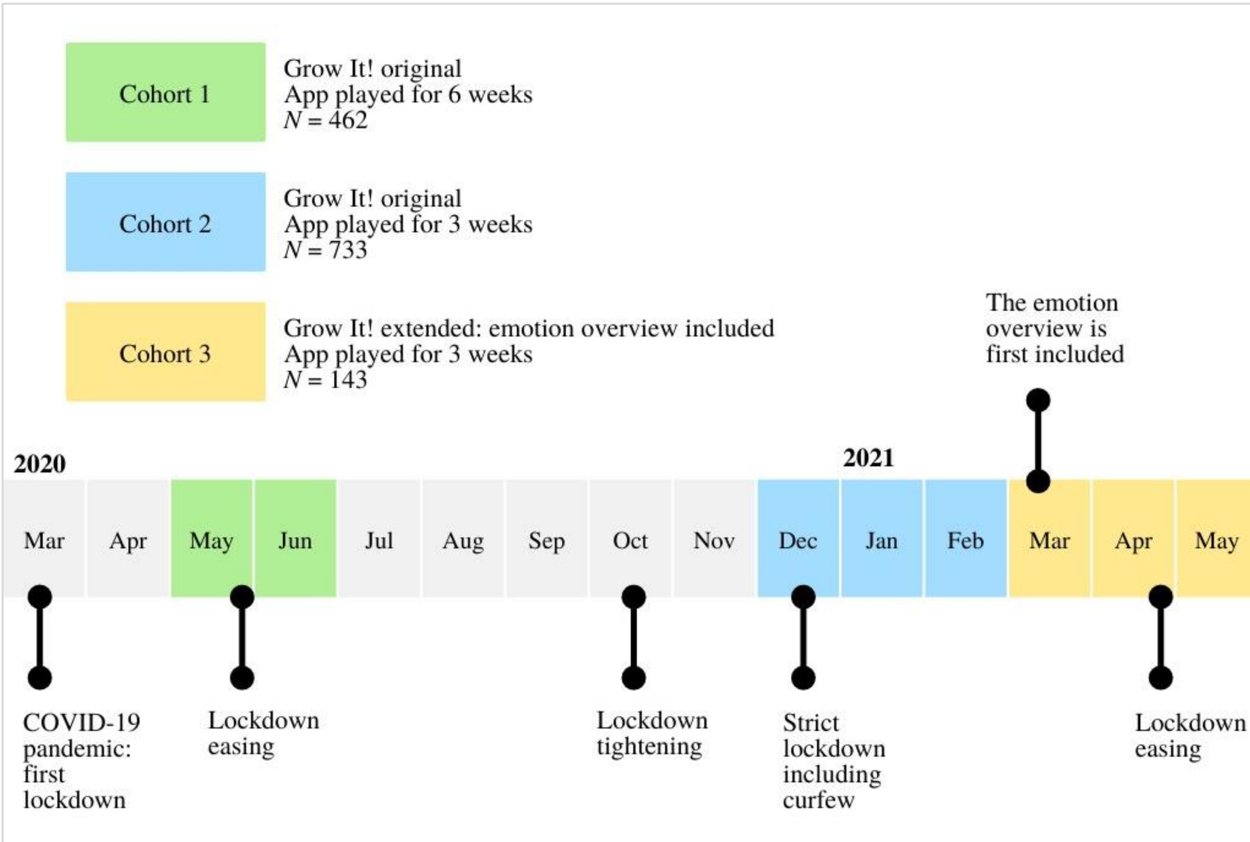
Methods

The Grow It! Corona Study

This study was a longitudinal online study regarding a multiplayer serious gaming app where (H_{1A} and H_{1B}) I examined changes in well-being within subjects. In addition, (H_2) I examined player activity between cohorts (i.e., app-versions). This study is part of a larger research project with multiple authors investigating Grow It!. During the COVID-19 pandemic, Grow It! was studied in three separate cohorts. The complete developmental process and current findings from cohort one and two are published elsewhere (Dietvorst et al., 2022a; Dietvorst et al., 2022b). I did not obtain the raw data for these cohorts. For the current study, we recruited a third cohort. Figure 2 illustrates the differences between the three cohorts. Online sources provide additional information (*Grow It! Corona project*, 2021). The app is furthermore being studied for other adolescent populations who are also at high risk for developing psychopathology. This includes adolescents that are chronically ill (chronic somatic conditions) and offspring of parents with psychiatric disorders.

Figure 2

Cohort Differences



Note. Illustrated above is a timeline that demonstrates at what timepoint registrations were open for cohort one (green), cohort two (blue), and cohort three (yellow), as well as their respective characteristic differences. Further details regarding the specific lockdown regulations can be found at https://www.rivm.nl/gedragsonderzoek/tijdlijn-maatregelen-covid.

Participants

We recruited Dutch adolescents aged 12 to 25 during the COVID-19 pandemic. Adolescents going through a pandemic are classified as at-risk adolescents due to the stressful societal circumstance and are therefore suitable candidates that may benefit from playing Grow It!. The total analytic sample size for the current study was 143 participants[1]. The majority of participants were female (71.3%), highly educated (75,5%) and held Dutch nationality (92.3%; see Table 1 for demographics). The mean age of participants

[1] As I wrote this paper for a master thesis, I received 40% of the data (randomly selected) from a larger dataset to analyze.

was 15.4 (*SD* = 3.2) which is younger than participants in cohort one (*M* = 16.7, *SD* = 3.4) and cohort two (*M* = 18.7, *SD* = 3.7; Dietvorst et al., 2022a). Appendix B illustrates a distribution of age and gender. We recruited participants through (social) media as well as their schools via their teacher's toolbox counseling lessons. Adolescents of the third cohort were able to enroll themselves into the study between the period of March 2021 until May 2021. Admissions were on a rolling basis meaning that they could start as soon as they had registered. Participation was possible for adolescents who signed the informed consent. For adolescents under 16 required an additional informed consent from the parents. We excluded anyone from participation who did not sufficiently comprehend and write the Dutch language and anyone who was not residing in the Netherlands at the time of the study. Recruitment was in Dutch ensuring sufficient language proficiency of participants. Participants did not receive compensation for their participation but were able to win prizes such as EarPods and gift vouchers. The Medical Ethical Committee of the Erasmus Medical Centre (registration number: MEC2020-0287) approved the study and followed the guidelines as stated by the World Medical Association Declaration of Helsinki (World Medical Association, 2013).

Table 1

Demographic Characteristics of Cohort Three

Characteristic	Cohort 3	
	n	%[a]
Gender		
Male	39	27.3
Female	102	71.3
Cultural identity		
Dutch	132	92.3
Unspecified	11	7.7
Education level[b]		
High	108	75.5
Middle	21	14.7
Low	9	6.3

[a] Does not add up to 100 because of missing values.

[b] Low = (preparatory school for) technical and vocational training, labeled in the Dutch education system as VMBO, praktijkonderwijs and MBO; Middle = (preparatory school for) professional education, labeled in the Dutch education system as HAVO, HAVO-VWO and HBO; High = (preparatory school for) university, labeled in the Dutch education system as VWO, Gymnasium and WO.

Procedure

Adolescents registered on the study website (www.growitapp.nl). After signing the informed consent on a secure webpage, the participants filled in the online questionnaire (the baseline questionnaire).

The questionnaire was self-administered questionnaire and concerned multiple choice and open questions about participants' demographic characteristics, well-being, depressive symptoms, anxiety, loneliness, psychological care, COVID-19 specific items and coping. The questionnaire took approximately eight minutes to complete. Once adolescents completed the online questionnaire an SMS was sent to the participants with a unique code for the app, which they could use to log in.

The adolescents used Grow It! over the course of a couple of weeks (Cohort one for six weeks, cohort two and three for three weeks). In the app adolescents were assigned to teams who competed against each other to grow a virtual tree (See Appendix C). To grow this tree, participants needed to earn points. Points could be earned in two ways. 1) Each day, at five random time intervals between 9:00 and 21:00, the participants got a notification to fill in the ESM questionnaire. Adolescents could record the following aspects: sleep, location, with whom, online contact, emotions, tired, loneliness, bored, worried, negative event, daily coping, worry related to COVID-19, atmosphere at home, exercise, positive event, and challenge rating. An ESM questionnaire took one to two minutes to fill in for which participants had a time window of 45 minutes. Adolescents received no more than two reminders per questionnaire. Cohort three was also able to view the results of the ESM questionnaires, displayed in emotion charts (see Figure 1). The emotion chart displayed each emotion in a bubble that varied in size depending on severity of the emotion. Charts also varied depending on the selected situation (e.g., with friends, at school, at home etc.). 2) The second way to earn points was to select and complete one challenge a day (see Appendix C). The rest of the time participants were free to interact with the app whenever they wanted. Teammates could for example interact with each other with positive stickers in the chat function (see Appendix C).

After three to six weeks in the app, participants received reminders via SMS to fill in the online questionnaire again as a follow-up, with an additional extensive user evaluation of the app. When participants had completed sufficient (more than half) questionnaires they were permitted to enter the prize raffle. Appendix D shows an illustration of the study timeline.

Measures

App Activity

The construct 'app activity' refers to the percentage of ESM questionnaires that were filled in (% ESM questionnaires = number of filled in ESM questionnaires / maximum number of ESM questionnaires). This is also called ESM compliance. I developed this measure myself. The ESM questionnaire is a short, self-administered questionnaire, commonly used in ESM research (Van Roekel et al., 2019). The ESM has high ecological validity as the questionnaires measure how a person feels in the moment and prevents recall bias (Larson & Csikszentmihalyi, 2014). The psychometric properties of the ESM questionnaire in Grow It! are unknown, however.

Affective Well-being

The construct ‘affective well-being’ was operationalized with the question “How happy did you feel last week?” with the answer provided on a 7-point Likert scale ranging from “Not at all” (1) to “Totally” (7) (Office for National Statistics, 2018; Tay, 2018). This construct was part of the baseline and follow-up online questionnaire. It has good convergent validity and strongly correlates with positive affect (Beyens et al., 2020).

Cognitive Well-being

The construct “cognitive well-being” was operationalized with the question “How satisfied did you feel with your life last week?” with the answer provided on a 10-point[2] Likert scale ranging from “Not at all” to “Totally” (Office for National Statistics, 2018; Tay, 2018). This construct was part of the baseline and follow-up online questionnaire. It demonstrates a substantial degree of criterion validity and has strong reliabilities, similar to that of multiple-item scales (Cheung & Lucas, 2014). Furthermore, evidence from Lucas et al. (1996) shows that affective and cognitive well-being are two distinct aspects of subjective well-being, demonstrating good divergent validity between the two constructs.

Emotion Overview Evaluation

We developed the ‘user experience’ measure as part of the online questionnaire by asking questions about various aspects of the app such as the ESM questionnaires, the challenges, the chat function, and the emotion chart. Users could evaluate the emotion chart with four questions. Question one was: “How clear did you find the emotion overview?” where answers included: “Not clear at all”, “A little bit clear”, “Pretty clear”, “Very clear”, or “Totally clear”. The second question: “How did the emotion overview affect you?” was answerable with: “I got to know myself better”, “It made me feel better”, “It made me think about how I feel more often”, “No effect”, or “Otherwise”. The third question was a yes-no question: “Did you share

---

[2] The reason that affective and cognitive well-being were rated on different scales is because 1) well-being constructs were replicated from Dietvorst et al. (2022a) and 2) the office for national statistics reports that people tend to answer more extremely on a “happiness yesterday” question than on question regarding “life satisfaction”. The reason may be that emotions shift more drastically over the days and hours, while life satisfaction is a more cognitive assessment on life overall (Office for National Statistics, 2011).

your emotion overview with others?", while the fourth question was an open question: "Who did you share the emotion overview with?".

Data analysis

The data analyses followed the pre-registered analysis plan (see Appendix A) and were executed using SPSS (IBM Corp., 2017). The syntax can be found in Appendix E. I calculated The ANOVA manually with an online calculator (https://statpages.info/anova1sm.html) as only summary data was available (see Appendix F). As pre-registered, the data I excluded from the analyses were any participants who did not have any scores for well-being. After inspection of the data, I decided to also exclude any participants who did not show any app activity (missing values for ESM compliance) and participants who did not (sufficiently) fill in the follow-up from the analyses. I defined a sufficient follow-up as participants who (at least) filled in both well-being questions and gave Grow It! extended a grade.

Prior to conducting the first analysis, I checked the following assumptions: (1) The assumption of independent subjects; I reasonably assumed that the paired observations (pre and post-test well-being scores) are independent between participants as they participated based on their individual interest and observations were collected independently from each participant. I assessed (2) the assumption of normally distributed differences by inspecting a histogram, QQ-plot, skewness, and kurtosis values of affective and cognitive well-being difference scores. Cognitive well-being differences were slightly positively skewed, but this was considered negligible due to small value for absolute skewness (.26). The results of these tests therefore confirmed the assumption of normality. Furthermore, formal tests such as the Shapiro-Wilk and Kolmogorov-Smirnov tests of normality showed non-normality. Notably, these tests are sensitive to minor deviations from normality and given that t-tests are robust against these deviations (Wilcox, 2011), I did not take results of formal testing into account. I checked (3) the assumption of no extreme outliers by inspecting boxplots of affective and cognitive well-being difference scores. I found no extreme outliers for affective well-being differences, however, concerning cognitive well-being two outliers deviated extremely. I reasonably assumed that running the analysis with extreme outliers included did not substantially impact the results, as extreme outliers have a relatively smaller impact on results in larger sample sizes (Van Selst & Jolicoeur 1994). In addition, outliers can potentially provide valuable insights, therefore they were kept in the analysis. For further inspection of assumptions see Appendix G.

I analyzed the first hypotheses (RO_1, $H_{1A,1B}$) with two one-tailed paired samples t-tests. To confirm the hypothesis that there is a significant increase in mean scores of affective and cognitive well-being from baseline to follow up, a t-score with a *p*-value smaller than 2.5% ($p \leq .025$) had to be found, rejecting the null hypothesis of no effect. This *p*-value was adjusted due to correction for multiple testing using the

Bonferroni method (Armstrong, 2014). Furthermore, I carried out an exploratory analysis where I investigated person-centered effect sizes to identify how many individuals meaningfully changed in their well-being (Grice et al., 2020), as pre-registered. To determine the threshold for a meaningful change we used a distribution-based approach. This approach considers a change of 0.2 standard deviations (small effect size according to Cohen's *d*) a threshold to identify meaningful changes beyond measurement error and random variation (Engel et al., 2018; Cohen, 1988).

Prior to conducting the second analysis, I checked the following assumptions: (1) observations are independent. (2) For the assumption of normally distributed data I visually inspected a histogram and QQ-plot and also assessed skewness, and kurtosis values of ESM compliance percentages. The histogram indicated slight positive skew of moderate strength (.52), however, this deviation from normality is not considered substantial. These results confirmed the assumption of normality for ESM compliance. A deviation from the pre-registration concerns the assumption "homogeneity of group variances." I initially intended to check this assumption; however, I did not get access to the data for cohort one and two rendering a Levene's test impossible to conduct. Nevertheless, I decided to proceed with the analysis (see Appendix G for inspection of assumptions).

I analyzed the second hypothesis (RO_2, H_2) with a one-way ANOVA. I used the mean, standard deviation, and sample size of the ESM compliance variable from cohort one and two using Grow It! original as reported in the paper by Dietvorst et al. (2022a) together with the raw data found for cohort three using Grow It! extended. To confirm the hypothesis that users of Grow It! extended will show more app activity than users of Grow It! original, an *F*-score with a *p*-value smaller than 5% ($p \leq .05$) had to be found, and the Post Hoc tests were required to show significant ($p \leq .05$) mean differences between cohort one and three as well as cohort two and three. I performed an additional effect size calculation for the ANOVA (not pre-registered). I used Cohen's *d* to assess the size of the effect between Grow It! original and Grow It! extended users, for which I pooled cohort one and two into one group (representing Grow It! original users). The calculations of the ANOVA and the effect size can be found in Appendix F.

I carried out an additional exploratory analysis that was not pre-registered to investigate whether there were any noteworthy correlations present in this rich data set of the third cohort. I did this by investigating Pearson correlations between seven continuous variables. The strength of the correlations was interpreted with Cohen's guidelines where correlations of .10, .30 and .50 are considered small, medium, and large, respectively (Cohen, 1988). I corrected for multiple comparisons with the Bonferroni method. As 21 correlations were compared, they were considered significant with a *p*-value smaller than 0.24% ($p \leq .002$). The other research objectives are of descriptive nature and do not require an analysis plan.

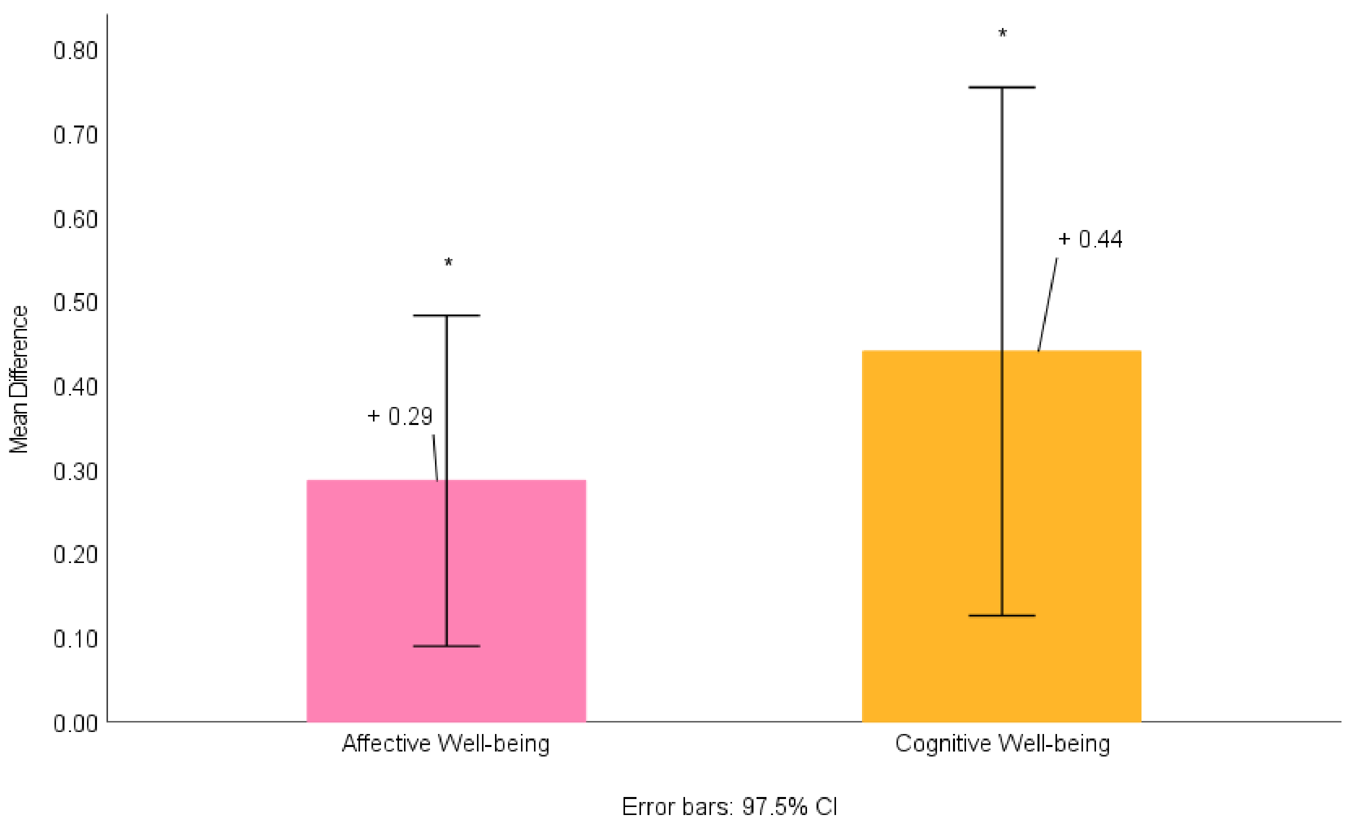
Results

RO$_1$: Differences in Well-being After Playing Grow It! Extended

To test the first hypothesis ($H_{1A,\ 1B}$: Adolescents will show an increase in their affective and cognitive well-being after playing Grow It! extended.), I conducted two paired samples t-tests comparing the baseline and follow-up scores of affective and cognitive well-being, respectively. Affective well-being was significantly higher at follow-up than at baseline ($M_{diff} = 0.29$, $SD_{diff} = 1.04$), $t(142) = 3.30$, $p < .001$ (one-tailed), $d = .28$. Cognitive well-being was also significantly higher at follow-up than at baseline ($M_{diff} = 0.44$, $SD_{diff} = 1.66$), $t(142) = 3.17$, $p < .001$ (one-tailed), $d = .27$. Figure 3 illustrates the mean increases in well-being.

Figure 3

Increases in Affective and Cognitive Well-being



Note. Illustrated in the bar chart is the mean increase in affective well-being (pink), and the mean increase in cognitive well-being (orange).The confidence intervals do not cross zero, demonstrating the significance of the results.

In addition, I conducted an exploratory analysis ($RO_{1, SUB}$) to identify how many individuals saw a difference in their well-being. Regarding affective well-being of the adolescents, I found that for 40% of participants the effect size of improvement met or exceeded the a priori criterion for a practically significant effect ($d \geq .20$). For 18% of participants the effect size of worsening met or exceeded the a priori criterion ($d \leq -.20$). For 42% of participants no practically significant effect was found ($-.20 < d < .20$). Regarding participants' cognitive well-being, 49% of participants saw an improvement ($d \geq .20$) in cognitive well-being, while 22% saw a worsening ($d \leq -.20$) and 29% remained the same ($-.20 < d < .20$). Taken together, the results of the first research objective compare similarly to cohort one and two (Dietvorst et al., 2022a). I confirmed hypotheses 1A and 1B, meaning that on average the participants showed meaningful increases in affective and cognitive well-being after playing Grow It! extended. The person-centered effect sizes (Grice et al., 2020) also provided evidence for well-being improvements. These findings collectively demonstrate positive differences in well-being on both the group and individual level.

RO_2: Differences in App Activity Between Grow It! Original and Grow It! Extended Users

To test the second hypothesis (H_2: Users of Grow It! extended will show more app activity than users of Grow It! original.), I conducted a one-way analysis of variance and showed that there was a significant difference in ESM compliance between all three cohorts $F(2,1335) = 53.13$, $p < .001$. In cohort one ($N = 462$) participants filled in an average of 14.2% of all ESM questionnaires ($SD = 20.3$), cohort two ($N = 733$) filled in an average of 20.6% ($SD = 25.8$) and cohort three ($N = 143$) filled in more than one third of the ESM questionnaires ($M = 37.6$, $SD = 23.7$). The post hoc analysis using Tukey's *HSD* revealed that between all cohorts there were significant mean differences, demonstrating that the amount of app activity (i.e., the completion of ESM questionnaires) meaningfully differed for each cohort. The size of the effect between Grow It! original users (cohort one and two) and Grow It! extended users (cohort three) was a large effect ($d = .80$). Due to significant post hoc comparisons between cohort three and one, as well as cohort three and two, I was able to confirm the second hypothesis.

RO_3: Evaluation of Grow It! Extended

The third research objective (RO_3) of this study was to explore how participants evaluated Grow It! extended. After inspection of user experience questions, I found that participants positively evaluated Grow It! extended as they gave it an average grade of 7.2 out of 10 ($SD = 2.0$). The average grade for the design of the app was 8.1 out of 10 ($SD = 1.3$). Out of all the participants 66.4% (95/143) would recommend the app to friends. These results are very similar to those in cohort one and two.

In addition, I inspected the evaluation of the emotion overview in further detail. All participants ($n = 94$, 65.8%) who self-reported on the effects of using the emotion overview report having experienced positive effects, while no-one reported a lack thereof or negative effects. Additionally, participants evaluated the emotion overview as sufficiently clear given that 67.2% of them rated the overview as "pretty clear" and above ($M = 3.3$, $SD = 1.2$). Some of the participants were comfortable enough to share the emotion overview with others: 12.6% reported sharing it. Those who specified shared it either with friends and/or family. See Table 2 for specifics regarding the user evaluation.

Table 2

Results Regarding the Evaluation of the Emotion Overview

User evaluation item	Cohort 3 ($N = 143$)	
	N	%[a]
Clarity of the emotion overview		
Not clear at all	7	4.9
A little bit clear	34	23.8
Pretty clear	31	21.7
Very clear	41	28.7
Totally clear	24	16.8
Self-reported effect of the emotion overview		
I got to know myself better	32	22.4
It made me feel better	13	9.1
It made me think about how I feel more often	49	34.3
No effect	0	0.0
Shared the emotion overview with someone		
Yes	18	12.6
With Family	8	5.6
With Friends	9	6.3
No	119	83.2

[a] Does not add up to 100 because of missing values.

Exploratory Pearson Correlations

I carried out an exploratory analysis of Pearson correlations with the data of cohort three, as demonstrated in Table 3. Results show a large correlation between the grade of the app and the grade of the design, which indicates that positive app evaluations are correlated. A medium strength correlation was found between the mean differences of affective and cognitive well-being. This indicates that improvements in well-being are positively related to each other. Notably, improvements in affective nor cognitive well-being did not correlate with ESM compliance or any other variables, suggesting that app activity may not be associated with well-being changes. Furthermore, results also indicate the presence of small strength correlations. The evaluative question regarding the clarity of the emotion overview positively correlated with ESM compliance. This indicates that the amount of activity in the app is related to how clear the adolescents found the emotion overview. Finally, a second small correlation was found between the clarity of the emotion overview and the grade Grow It! was given, which indicates that there is a positive association between the comprehension of the emotion overview and the rating of the app.

Table 3

Exploratory Pearson Correlations in Cohort Three

Variable	A	B	C	D	E	F	G
A. Age	1						
B. ESM compliance	.12	1					
C. Grade design	**-.17**†	**.24**†	1				
D. Grade Grow It!	-.13	**.22**†	**.54***	1			
E. Clarity EO	**-.17**†	**.29***	.16	**.28***	1		
F. M_{diff} Affective WB	.06	-.011	-.05	-.11	-.05	1	
G. M_{diff} Cognitive WB	-.02	.04	-.13	-.01	-.07	**.47***	1

† Correlation is significant at the .05 level (uncorrected alpha)

* Correlation is significant at the .002 level (2-tailed; corrected alpha)

EO = emotion overview, M_{diff} = Mean difference, WB = well-being

Discussion

Studying the effects of m-health apps like Grow It! is crucial for understanding how to provide accessible and easy to use (mental) healthcare to people at high risk for developing psychopathology, such as adolescents during the COVID-19 pandemic. To understand the underlying mechanisms of what makes m-health effective, it is necessary to investigate individual elements such as real-time personalized feedback. In the current study I aimed to replicate improvements in adolescents affective and cognitive well-being after using Grow It!, as initially found by Dietvorst et al. (2022a). I successfully replicated these results on both a group and individual level, using the extended version of Grow It! which included the emotion overview feedback element. Secondly, I investigated whether users who play Grow It! extended would complete more ESM questionnaires than users who played Grow It! original. This study found that adolescents engaged significantly more with ESM questionnaires in Grow It! extended. Thirdly, I aimed to describe how adolescents evaluated the added emotion overview feature. Similar to evaluations in Dietvorst et al. (2022b) users still evaluated the app positively. Adolescents welcomed the addition of the emotion overview as they self-reported positive effects from using it and desired to share it with friends and family.

Improvements in Well-being

Regarding the first hypothesis individual level improvements were the following: in cohort three 40% of participants saw a positive increase in their affective well-being, similar to cohort one (45%) and two (42%; Dietvorst et al. 2020a). Regarding cognitive well-being, 49% of participants increased in cohort three, 53% in cohort one and 45% in cohort two. In addition to individual level findings, small to medium effect sizes found on a group level complement the substantial improvements in well-being. Cohort one increased the most in affective well-being (+0.42, d = .32), followed by cohort three (+0.29, d = .28) and cohort two (+ 0.32, d = .23). Regarding cognitive well-being cohort one (+0.57, d = .27) and cohort three (+0.44, d = .27) demonstrated equal sized improvements, while cohort two (+0.43, d = .20) improved to a lesser extent. Possible explanations may relate to cohort characteristics. Cohort one played the app for double the amount of time, increasing intervention exposure. In addition, cohort three has higher engagement levels than other cohorts. Both cohort one and three improve more than cohort two, potentially indicating a positive dose-response relationship between exposure to Grow It! and well-being improvements. Including confidence intervals for effect size interpretation would be beneficial for adding robustness.

Notably, my exploratory correlation analysis, consistent with findings in Dietvorst et al. (2020a), did not find a significant correlation between well-being improvements and ESM compliance, suggesting that the observed increases in well-being may also be regression to the mean. If Grow It! had a causal effect on well-being one would expect that participants who are more compliant with filling in ESM questionnaires (i.e., treated more) should experience more therapeutic benefits than participants who do not (i.e., treated less). Potential reasons for the absence of this correlation may be that ESM compliance by itself may not accurately reflect the combined therapeutic effect of all the validated elements Grow It! consists of, such as CBT-based coping challenges, chat functions and other game related activities. I considered it more appropriate to exclusively use ESM compliance as a measure of app activity due to its sensitivity in capturing the effects of including the emotion overview, as it is formed directly from completed ESM questionnaires. A limitation of this study pertains to the absence of a control group, raising the question whether improvements in well-being are attributable to Grow It! or other factors. However, even after conducting a sensitivity analysis to investigate possible confounders such as education level, age, gender, and COVID-19 stringency, Dietvorst et al. (2022a) still found increases in well-being. Therefore, it is unlikely these factors played a significant role in the current study. Nevertheless, to be able to make any true claims on the causal relationship of the effect of Grow It! on well-being and whether the emotion overview has added any significant value to its effectiveness it is strongly suggested to conduct a randomized control trial (RCT). Bakker et al. (2016) also mention the need for RCT studies. An RCT was not feasible for the current study due to practical and ethical reasons aimed at ensuring equal access to Grow It! for all young adults during the COVID-19 pandemic. Researchers are addressing this limitation in future studies, as they currently are conducting a RCT study for Grow It! with multiple follow-ups.

Increases in App Activity

Findings regarding the second hypothesis show that Grow It! extended users fill in a meaningfully higher amount of ESM questionnaires than for Grow It! original users. This indicates that the addition of the emotion overview appears to have a positive effect on interaction with the app, supporting gamification and theories that state that feedback increases engagement (Cugelman, 2013; Robson et al., 2015). However, I identified several challenges that complicate the interpretation of these findings. Firstly, the homogeneity of variances assumption was not met due to inaccessible data of cohort one and two. The validity of the results is at risk, increasing the likelihood that any conclusions drawn may be false. Secondly, the three cohorts analyzed may not be entirely comparable. Cohort one played the app twice as long (six weeks) as cohort two and three (three weeks). A natural decline in compliance could perhaps explain why this study found a significant difference between cohort one and two. This same reasoning could hypothetically explain the significant difference between cohort three and one, however it does not account for the significant effect between cohort three and two as they played the app for equal amounts of time. It is therefore possible that the included emotion overview may account for significant difference in app activity. These considerations demonstrate the need for further investigation, as multiple factors may be at play.

App Evaluations

The user evaluations show that Grow It! adolescents rated the app very similarly to Dietvorst et al., (2022b), with cohort two and three rating it a 7.2, while cohort one rated the app a 7.1 out of 10. Cohort three demonstrates the highest evaluation in design (8.1), followed by cohort two (8) and cohort one (7.7) and are likely not significantly different. The app is still being recommended by 66.4% teens, which is lower than in cohort one (72.6%) and two (75.6%). This might indicate that the addition of the emotion overview does not have effects on the general evaluation of the app. A different explanation may be that Grow It! was more appealing in the beginning of the pandemic, due to novelty and stricter lockdown measures. The design grade is higher; however, it is not clear to what extent this is the result of the visual appeal of the emotion overview. Nevertheless, the evaluation of the emotion overview is positive: individuals reported gaining more insight and knowledge on their emotions and desire to share the overview with friends and family, suggesting that the emotion overview may have been a valuable addition for the usability of Grow It!. Regarding the clarity of the emotion overview, most adolescents found it sufficiently clear, however there is a minority which reported otherwise. Exploratory correlations found that the clarity of the emotion overview impacts how positively teens rated Grow It!. In addition, it correlated with ESM compliance, indicating that clarity may play a substantial role in the level of engagement. Between the design grade and overall app grade there is also a positive association, indicating that a good design plays an important role for a high app rating. Therefore, adjusting the design to improve the clarity of the emotion overview is highly recommended and careful attention should be paid to ensure that children comprehend its purpose and functioning. These correlations further support that the emotion overview has an impact on app activity and evaluation.

Further Limitations and Future Directions

In this study I removed 56 participants (41.8% of initial dataset) from the analyses who failed to fill in most of the follow-up and/or did not have any values for ESM compliance. Dietvorst et al. (2022b) excluded participants who did not show app activity (i.e., completed zero or one activity of challenges or ESM) from their user-evaluation analysis. In cohort one, 302 participants (44.1%) did not fill in the follow-up, while in cohort two this was 391 participants (37.8%). Attrition rates look similar when across cohorts, though statistical confirmation may be necessary. Taking increased engagement levels in cohort three (H_2) into account, findings imply that those who were initially engaged became even more involved, likely because of the newly included feedback. Even though dropout rates were similar to other web-based studies (Dietvorst et al., 2020b), attrition bias is a considerable limitation. Arguably, the exclusion of participants may have led to a non-random sample which compromised the validity of the results, causing misrepresentation of app experiences and participant characteristics. Nevertheless, the justification for removing these participants is that their lack of engagement undermines the true representation of the effects of Grow It! extended. A lack of follow-up is insufficient to draw valuable conclusions from regarding app functionality.

The future directions for Grow It! include additional research with enhanced research designs to improve reliability and validity. Future research should include various adolescent populations with varying treatment motivations to assess the applicability of Grow It!. Furthermore, it would be valuable to investigate whether improvements in well-being last over the long-term. Based on this current study I suggest creating a more accurate measure of app activity to help better evaluate app effectiveness. It is also worth considering that targeting ESM compliance may not be the best strategy for improving Grow It! as my findings demonstrate that although including the emotion overview increased app activity, this does not necessarily guarantee a better therapeutic outcome. Instead, future research might benefit from targeting lower attrition rates and finetuning the app. Examples include better tailoring to adolescents (e.g., age-appropriate design and rewards), modifying dosage (e.g., less frequent ESM questionnaires) and enhancing immersion (e.g., by incorporating a storyline). Bakker et al. (2016) furthermore suggests providing adolescents with mental health information and links to crisis support services. Interviews with doctors could be another useful step in improving Grow It!. as they understand the needs and capabilities of their patients. Their expertise can inform effective m-health and how it should be implemented in clinical practice. Additionally, artificial intelligence (AI) techniques such as machine learning could be a brilliant addition to Grow It! for early detection and prevention of psychopathology, by signaling clinicians and tailoring the intervention more accurately to its user (Lovejoy, 2019), supporting a broader scope of adolescents beyond those already affected by mental illness.

Conclusion

Taken together, including the emotion overview feature emerges as a useful addition to Grow It! as app engagement levels surpass previous records and user evaluations signify a notable level of enthusiasm among adolescents towards the app. Grow It! encompasses a multitude of essential components that contribute to the formulation of an effective m-health app. It proves to be user-friendly, and its strength lies in bringing simple yet evidence-based CBT techniques in an entertaining way. It has the potential to normalize taking care of one's mental health and lowers the hurdle to seek help. Although improvements in well-being were replicated, based on these findings alone it remains uncertain whether real-time personalized feedback is an element that underpins the effectiveness of the app. As a result, further research with a RCT research design is highly recommended to finetune and prove the effectiveness of Grow It! and further explore what role feedback plays. With an ongoing RCT study currently underway, it will further contribute to and bring Grow It! to its full potential in supporting at-risk adolescents with their mental health and create a happier and healthier generation in the long run.

References

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*(5), 502–508. https://doi.org/10.1111/opo.12131

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental health smartphone apps: Review and evidence-based recommendations for future developments. *JMIR Mental Health*, *3*(1), e7. https://doi.org/10.2196/mental.4984

Bastiaansen, J. A., Meurs, M., Stelwagen, R., Wunderink, L., Schoevers, R. A., Wichers, M., & Oldehinkel, A. J. (2018). Self-monitoring and personalized feedback based on the experiencing sampling method as a tool to boost depression treatment: A protocol of a pragmatic randomized controlled trial (ZELF-i). *BMC Psychiatry*, *18*(1), 276. https://doi.org/10.1186/s12888-018-1847-z

Bergin, A. D., Vallejos, E. P., Davies, E. B., Daley, D., Ford, T., Harold, G., Hetrick, S., Kidner, M., Long, Y., Merry, S., Morriss, R., Sayal, K., Sonuga-Barke, E., Robinson, J., Torous, J., & Hollis, C. (2020). Preventive digital mental health interventions for children and young people: A review of the design and reporting of research. *Npj Digital Medicine*, *3*(1), 133. https://doi.org/10.1038/s41746-020-00339-7

Beyens, I., Pouwels, J. L., Driel, I. I. van, Keijsers, L., & Valkenburg, P. M. (2020). The effect of social media on well-being differs from adolescent to adolescent. *Scientific Reports*, *10*(1), 10763. https://doi.org/10.1038/s41598-020-67727-7

Bobek, E., & Tversky, B. (2016). Creating visual explanations improves learning. *Cognitive Research: Principles and Implications*, *1*, 27. https://doi.org/10.1186/s41235-016-0031-6

Burgos, D., Nimwegen, C., Oostendorp, H., & Koper, R. (2007). Game-based learning and the role of feedback: A case study. *Advanced Technology for Learning*, *4*. https://doi.org/10.2316/Journal.208.2007.4.208-0918

Cheung, F., & Lucas, R. E. (2014). Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Quality of Life Research*, *23*(10), 2809–2818. https://doi.org/10.1007/s11136-014-0726-4

Clarke, A. M., Kuosmanen, T., & Barry, M. M. (2015). A systematic review of online youth mental health promotion and prevention interventions. *Journal of Youth and Adolescence*, *44*(1), 90–113. https://doi.org/10.1007/s10964-014-0165-0

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. https://www.utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf

Cowley, B., Charles, D., Black, M., & Hickey, R. (2008). Toward an understanding of flow in video games. *Computers in Entertainment*, *6*(2), 20:1-27. https://doi.org/10.1145/1371216.1371223

Csikszentmihalyi M. (1990). *Flow: The psychology of optimal experience*. Harper & Row. https://www.researchgate.net/publication/224927532_Flow_The_Psychology_of_Optimal_Experience

Cugelman, B. (2013). Gamification: What it is and why it matters to digital health behavior change developers. *JMIR Serious Games*, *1*(1), e3139. https://doi.org/10.2196/games.3139

Dietvorst, E., Aukes, M. A., Legerstee, J. S., Vreeker, A., Hrehovcsik, M. M., Keijsers, L., & Hillegers, M. H. J. (2022b). A smartphone serious game for adolescents (Grow It! app): Development, feasibility, and acceptance study. *JMIR Formative Research*, *6*(3), e29832. https://doi.org/10.2196/29832

Dietvorst, E., Legerstee, J. S., Vreeker, A., Koval, S., Mens, M. M., Keijsers, L., & Hillegers, M. H. J. (2022a). The Grow It! app—Longitudinal changes in adolescent well-being during the COVID-19 pandemic: A proof-of-concept study. *European Child & Adolescent Psychiatry*. *32*, 1097–1107. https://doi.org/10.1007/s00787-022-01982-z

Engel, L., Beaton, D. E., & Touma, Z. (2018). Minimal clinically important difference. *Rheumatic Disease Clinics of North America*, *44*(2), 177–188. https://doi.org/10.1016/j.rdc.2018.01.011

Free, C., Phillips, G., Felix, L., Galli, L., Patel, V., & Edwards, P. (2010). The effectiveness of m-health technologies for improving health and health services: A systematic review protocol. *BMC Research Notes*, *3*(1), 250. https://doi.org/10.1186/1756-0500-3-250

Friedler, B., Crapser, J., & McCullough, L. (2015). One is the deadliest number: The detrimental effects of social isolation on cerebrovascular diseases and cognition. *Acta Neuropathologica*, *129*(4), 493–509. https://doi.org/10.1007/s00401-014-1377-9

Gorrese, A., & Ruggieri, R. (2012). Peer attachment: A meta-analytic review of gender and age differences and associations with parent attachment. *Journal of Youth and Adolescence*, *41*(5), 650–672. https://doi.org/10.1007/s10964-012-9759-6

Grice, J. W., Medellin, E., Jones, I., Horvath, S., McDaniel, H., O'lansen, C., & Baker, M. (2020). Persons as effect sizes. *Advances in Methods and Practices in Psychological Science*, *3*(4), 443–455. https://doi.org/10.1177/2515245920922982

Grow It! Corona project. (2021). OSF. https://osf.io/2at58/?view_only=b691104ecc3d45ad8b48e1bd60ad7125

IBM Corp. (2017). *IBM SPSS Statistics for Windows*. IBM Corp.

Janse, P. D., de Jong, K., Veerkamp, C., van Dijk, M. K., Hutschemaekers, G. J. M., & Verbraak, M. J. P. M. (2020). The effect of feedback-informed cognitive behavioral therapy on treatment outcome: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *88*(9), 818–828. https://doi.org/10.1037/ccp0000549

Kramer, I., Simons, C. J. P., Hartmann, J. A., Menne-Lothmann, C., Viechtbauer, W., Peeters, F., Schruers, K., van Bemmel, A. L., Myin-Germeys, I., Delespaul, P., van Os, J., & Wichers, M. (2014). A therapeutic application of the experience sampling method in the treatment of depression: A randomized controlled trial. *World Psychiatry*, *13*(1), 68–77. https://doi.org/10.1002/wps.20090

Laamarti, F., Eid, M., & El Saddik, A. (2014). An overview of serious games. *International Journal of Computer Games Technology*, *2014*. https://doi.org/10.1155/2014/358152

Larson, R., & Csikszentmihalyi, M. (2014). The experience sampling method. In: *Flow and the foundations of positive psychology: The collected works of Mihaly Csikszentmihalyi* (pp. 21–34). Springer Netherlands. https://doi.org/10.1007/978-94-017-9088-8_2

Lovejoy, C. A. (2019). Technology and mental health: The role of artificial intelligence. *European Psychiatry*, *55*, 1–3. https://doi.org/10.1016/j.eurpsy.2018.08.004

Lucas, R. E., Diener, E., & Suh, E. (1996). Discriminant validity of well-being measures. *Journal of Personality and Social Psychology*, *71*(3), 616–628. https://doi.org/10.1037/0022-3514.71.3.616

March-Llanes, J., Marqués-Feixa, L., Mezquita, L., Fañanás, L., & Moya-Higueras, J. (2017). Stressful life events during adolescence and risk for externalizing and internalizing psychopathology: A meta-analysis. *European Child & Adolescent Psychiatry*, *26*(12), 1409–1422. https://doi.org/10.1007/s00787-017-0996-9

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry*, *17*(2), 123–132. https://doi.org/10.1002/wps.20513

Office for National Statistics. (2011). *Initial investigation into subjective well-being from the opinions survey*. National Archives UK. https://webarchive.nationalarchives.gov.uk/ukgwa/20160107223320/http:/www.ons.gov.uk/ons/rel/wellbeing/measuring-subjective-wellbeing-in-the-uk/investigation-of-subjective-well-being-data-from-the-ons-opinions-survey/initial-investigation-into-subjective-well-being-from-the-opinions-survey.html

Office for National Statistics. (2018). *Personal well-being user guidance.* Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/personalwellbeingsurveyuserguide.

Robson, K., Plangger, K., Kietzmann, J. H., McCarthy, I., & Pitt, L. (2015). Is it all a game? Understanding the principles of gamification. *Business Horizons*, *58*(4), 411–420. https://doi.org/10.1016/j.bushor.2015.03.006

Tay, L. (2018). *Handbook of well-being* (E. Diener, & S. Oishi Eds.). DEF Publishers. https://www.nobascholar.com/books/1

Van Roekel, E., Keijsers, L., & Chung, J. M. (2019). A review of current ambulatory assessment studies in adolescent samples and practical recommendations. *Journal of Research on Adolescence*, *29*(3), 560–577. https://doi.org/10.1111/jora.12471

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology Section A*, *47*(3), 631–650. https://doi.org/10.1080/14640749408401131

Wilcox, R. R. (2011). *Introduction to robust estimation and hypothesis testing*. Academic Press. https://books.google.nl/books?id=8f8nBb4__EYC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, *310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Appendix A

Pre-registration

Study Information

1. Title
   1.1. Provide the working title of your study. It may be the same title that you submit for publication of your final manuscript, but it is not a requirement.

The Grow It! App: Examining the Role of Real-time Personalised Feedback

2. Authorship

Louise Meyser

3. Research Questions
   3.1. Please list each research question included in this study.

RQ1: How does affective and cognitive well-being change after playing Grow it! extended?

RQ2: How does app activity (i.e., compliance of Experience Sampling Method) differ in users of Grow It! extended versus Grow It! original?

4. Hypotheses
   4.1. For each of the research questions listed in the previous section, provide one or multiple specific and testable hypotheses. Please state if the hypotheses are directional or non-directional. If directional, state the direction. A predicted effect is also appropriate here.

RQ1: How does affective and cognitive well-being change after playing Grow it! extended?

H_{1A}: Adolescents will show an increase in their affective well-being after playing Grow It! extended.
H_{1B}: Adolescents will show an increase in their cognitive well-being after playing Grow It! extended.

RQ2: How does app activity (i.e., compliance of Experience Sampling Method) differ in users of Grow It! extended versus Grow It! original?

H_2: Users of Grow It! extended will show more app activity than users of Grow It! original.

Sampling Plan

In this section we will ask you to describe how you plan to collect samples, as well as the number of samples you plan to collect and your rationale for this decision. Please keep in mind that the data described in this section should be the actual data used for analysis, so if you are using a subset of a larger dataset, please describe the subset that will actually be used in your study.

5. Existing data
    5.1. Preregistration is designed to make clear the distinction between confirmatory tests, specified prior to seeing the data, and exploratory analyses conducted after observing the data. Therefore, creating a research plan in which existing data will be used presents unique challenges. Please select the description that best describes your situation. Please do not hesitate to contact us if you have questions about how to answer this question (prereg@cos.io).
        5.1.1. Registration prior to creation of data: As of the date of submission of this research plan for preregistration, the data have not yet been collected, created, or realized.
        5.1.2. Registration prior to any human observation of the data: As of the date of submission, the data exist but have not yet been quantified, constructed, observed, or reported by anyone - including individuals that are not associated with the proposed study. Examples include museum specimens that have not been measured and data that have been collected by non-human collectors and are inaccessible.
        5.1.3. **Registration prior to accessing the data**: As of the date of submission, the data exist, but have not been accessed by you or your collaborators. Commonly, this includes data that has been collected by another researcher or institution.
        5.1.4. Registration prior to analysis of the data: As of the date of submission, the data exist and you have accessed it, though no analysis has been conducted related to the research plan (including calculation of summary statistics). A common situation for this scenario when a large dataset exists that is used for many different studies over time, or when a data set is randomly split into a sample for exploratory analyses, and the other section of data is reserved for later confirmatory data analysis.
        5.1.5. Registration following analysis of the data: As of the date of submission, you have accessed and analyzed some of the data relevant to the research plan. This includes preliminary analysis of variables, calculation of descriptive statistics, and observation of data distributions. Studies that fall into this category are ineligible for the Pre-Reg Challenge. Please contact us (prereg@cos.io) and we will be happy to help you.

6. Explanation of existing data
    6.1. If you indicate that you will be using some data that already exist in this study, please describe the steps you have taken to assure that you are unaware of any

patterns or summary statistics in the data. This may include an explanation of how access to the data has been limited, who has observed the data, or how you have avoided observing any analysis of the specific data you will use in your study. The purpose of this question is to assure that the line between confirmatory and exploratory analysis is clear.

For the current study, I will make use of data from 3 cohorts. Cohort 1 and 2 have previously been used for publications (Dietvorst et al., 2022a; Dietvorst et al., 2022b).

Data for cohort 3 has so far not been used for any publications. As an intern I do not have access to the data. Evelien Dietvorst (thesis supervisor Erasmus MC) has pre-processed and observed the data. I will be able to access the data once the pre-registration has been frozen.

Cohort 1
Cohort 1 was recruited between May 2020 and June 2020, during the first wave of the COVID-pandemic. Adolescents in cohort 1 played the app for 6 weeks. The sample consists of 1282 participants at baseline and 462 at follow up (Dietvorst et al. 2022a; Dietvorst et al. 2022b).

Cohort 2
Cohort 2 was recruited between December 2020 and March 2021, during the second wave of the COVID-19 pandemic. Adolescents in cohort 2 played the app for 3 weeks. The sample consists of 1871 participants at baseline and 733 at follow up (Dietvorst et al. 2022a; Dietvorst et al. 2022b).

Cohort 3
Cohort 3 was recruited between March 2021 and May 2021. The app for this samples includes the emotion overview. Adolescents in cohort 3 played the app for 3 weeks. The sample consists of roughly 400 participants.

All cohorts differ in terms of governmental restrictions that were in place at the different recruitment time periods.

7. Data collection procedures.
   7.1. Please describe the process by which you will collect your data. If you are using human subjects, this should include the population from which you obtain subjects, recruitment efforts, payment for participation, how subjects will be selected for eligibility from the initial pool (e.g. inclusion and exclusion rules), and your study timeline. For studies that don't include human subjects, include information about how you will collect samples, duration of data gathering efforts, source or location of samples, or batch numbers you will use.

The aim was to recruit Dutch adolescents of the ages 12 to 25, during the COVID-19 pandemic. Participants were recruited through social media as well as their schools via their teacher's toolbox counselling lessons. Adolescents were invited to enroll themselves into the study via the Grow It! website between the period of March 2021 until May 2021. Participation was possible for adolescents who signed the informed consent. For adolescents under 16 an additional informed consent from the parents was required. Anyone who does not sufficiently comprehend and write the Dutch language and anyone who is not residing in the Netherlands at the time of the study was excluded from participation.

8. Sample size
   8.1. Describe the sample size of your study. How many units will be analyzed in the study? This could be the number of people, birds, classrooms, plots, interactions, or countries included. If the units are not individuals, then describe the size requirements for each unit. If you are using a clustered or multilevel design, how many units are you collecting at each level of the analysis?

The aimed sample size for the Grow it! study was N=400 adolescents. As I may use 40% of the complete dataset for my master thesis, the estimated sample size for my study is around 160 participants.

9. Sample size rationale
   9.1. This could include a power analysis or an arbitrary constraint such as time, money, or personnel.

The reason for this sample size was constraint regarding COVID-19: data was only collected from March 2021 until May 2021 during the COVID-19 pandemic when restrictions were in place.

10. Stopping rule
    10.1. If your data collection procedures do not give you full control over your exact sample size, specify how you will decide when to terminate your data collection.

The data collection period took place during the third COVID-19 wave. May 31st was the last date that a participant could start in the app.

Variables

In this section you can describe all variables (both manipulated and measured variables) that will later be used in your confirmatory analysis plan. In your analysis plan, you will have the opportunity to describe how each variable will be used. If you have variables that you are measuring for exploratory analyses, you are not required to list them, though you are permitted to do so.

11. Manipulated variables
    11.1. Describe all variables you plan to manipulate and the levels or treatment arms of each variable. For observational studies and meta-analyses, simply state that this is not applicable.

The Grow It! app aims to increase affective and cognitive wellbeing of users. This has previously been shown in publications (Dietvorst et al., 2022a; Dietvorst et al., 2022b)

12. Measured variables
    12.1. Describe each variable that you will measure. This will include outcome measures, as well as any predictors or covariates that you will measure. You do not need to include any variables that you plan on collecting if they are not going to be included in the confirmatory analyses of this study.

The measured variables for the first hypothesis are **affective and cognitive well-being**. The construct ‘affective well-being’ is operationalized with the question “How happy did you feel last week?” with the answer provided on a 7-point Likert scale ranging from “Not at all” (1) to “Totally” (7). The construct ‘cognitive well-being’ is operationalized with the question “How satisfied did you feel with your life last week?” with the answer provided on a 10-point Likert scale ranging from “Not at all” to “Totally”. Affective and cognitive wellbeing are measures at baseline and at follow up (3 weeks thereafter).

App activity. App activity, also called ESM compliance refers to the percentage of completed ESM questionnaires.

13. Indices
    13.1. If any measurements are going to be combined into an index (or even a mean), what measures will you use and how will they be combined? Include either a formula or a precise description of your method. If you are using a more complicated statistical method to combine measures (e.g. a factor analysis), you can note that here but describe the exact method in the analysis plan section.

App activity is calculated with the following formula: % ESM questionnaires = number of filled in ESM questionnaires / maximum number of ESM questionnaires.

Design Plan

In this section, you will be asked to describe the overall design of your study. Remember that this research plan is designed to register a single study, so if you have multiple experimental designs, please complete a separate preregistration.

14. Study type
    14.1. **Observational Study** - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, natural experiments, and regression discontinuity designs.

15. Blinding
    15.1. Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.
        15.1.1. **No blinding is involved in this study**.

16. Study design
    16.1. Describe your study design. Examples include two-group, factorial, randomized block, and repeated measures. Is it a between (unpaired), within-subject (paired), or mixed design? Describe any counterbalancing required. Typical study designs for observation studies include cohort, cross sectional, and case-control studies.

This study is a longitudinal online mHealth study. Regarding the first hypothesis (A and B), a within-subjects design is used. Regarding the second hypothesis, the study uses a between-subject design between 3 cohorts.

17. Randomization
    17.1. If you are doing a randomized study, how will you randomize, and at what level

Does not apply.

Analysis Plan

You may describe one or more confirmatory analysis in this preregistration. Please remember that all analyses specified below must be reported in the final article, and any additional analyses must be noted as exploratory or hypothesis generating.

A confirmatory analysis plan must state up front which variables are predictors (independent) and which are the outcomes (dependent), otherwise it is an exploratory analysis. You are allowed to describe any exploratory work here, but a clear confirmatory analysis is required.

18. Statistical models
    18.1. What statistical model will you use to test each hypothesis? Please include the type of model (e.g. ANOVA, multiple regression, SEM, etc) and the specification of the model (this includes each variable that will be included as predictors,

outcomes, or covariates). Please specify any interactions that will be tested and remember that any test not included here must be noted as an exploratory test in your final article.

The first hypotheses (RO_1, $H_{1A,1B}$) will be analysed with two one-tailed paired samples t-tests ((Howell, 2013) ;Replication following Dietvorst et al., 2022a)

The mean affective well-being at baseline will be compared to the mean affective wellbeing at follow-up, 3 weeks after.
The mean cognitive well-being at baseline will be compared to the mean cognitive wellbeing at follow-up, 3 weeks after.

Hypothesis 2 will be analysed with a one-way ANOVA (3 group comparison; 1-2, 2-3, 1-3). Adolescents who play Grow It! extended (cohort 3) will be compared on their mean scores for ESM compliance to adolescents who play Grow It! original (cohort 1 and 2).

For the hypotheses the assumptions (independent samples, normally distributed data and absence of extreme outliers and homogeneity of group variances) will be checked and appropriate measures will be taken if they are not met (Pallant, 2016).

19. Transformations

19.1. If you plan on transforming, centering, recoding the data, or will require a coding scheme for categorical variables, please describe that process.

Does not apply.

20. Follow-up analyses

20.1. If not specified previously, will you be conducting any confirmatory analyses to follow up on effects in your statistical model, such as subgroup analyses, pairwise or complex contrasts, or follow-up tests from interactions. Remember that any analyses not specified in this research plan must be noted as exploratory.

Pairwise comparisons (Post-hoc) will be performed after the ANOVA to see which groups specifically show significant mean differences.

21. Inference criteria

21.1. What criteria will you use to make inferences? Please describe the information you will use (e.g. p-values, Bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?

To confirm the first hypotheses a t-score with a p-value smaller than 2,5% ($\alpha \leq 0{,}025$) has to be found, rejecting the null hypothesis of no effect. This p-value is adjusted due to correction for multiple testing using the Bonferroni method (Armstrong, 2014).

To confirm the second hypothesis an F-score with a p-value smaller than 5% ($\alpha \leq 0{,}05$) has to be found, and the Post Hoc tests show significant ($\alpha \leq 0{,}05$) mean differences between cohort 1 and 3 as well as cohort 2 and 3 (Howell, 2013)

22. Data exclusion
    22.1. How will you determine what data or samples, if any, to exclude from your analyses? How will outliers be handled?

I will inspect outliers and remove them if necessary. An elaborate description of outliers will follow in the manuscript of the thesis.

23. Missing data
    23.1. How will you deal with incomplete or missing data?

I will remove any participants with missing data regarding affective and cognitive wellbeing.

24. Exploratory analysis (optional)
    24.1. If you plan to explore your data set to look for unexpected differences or relationships, you may describe those tests here. An exploratory test is any test where a prediction is not made up front, or there are multiple possible tests that you are going to use. A statistically significant finding in an exploratory test is a great way to form a new confirmatory hypothesis, which could be registered at a later time.

Following H1, I aim to describe in percentages how many adolescents improve in their well-being. Grice effect sizes will be used to interpret how meaningful the change is. Anything above 0,2SD (small effect) will be considered as a meaningful increase.

Furthermore, I aim to describe results from the **user evaluation** regarding the emotion overview, to understand how adolescents evaluate this addition.

Script (Optional)

The purpose of a fully commented analysis script is to unambiguously provide the responses to all of the questions raised in the analysis section. This step is not common, but we encourage you to try to create an analysis script, refine it using a modeled dataset, and use it in place of your written analysis plan.

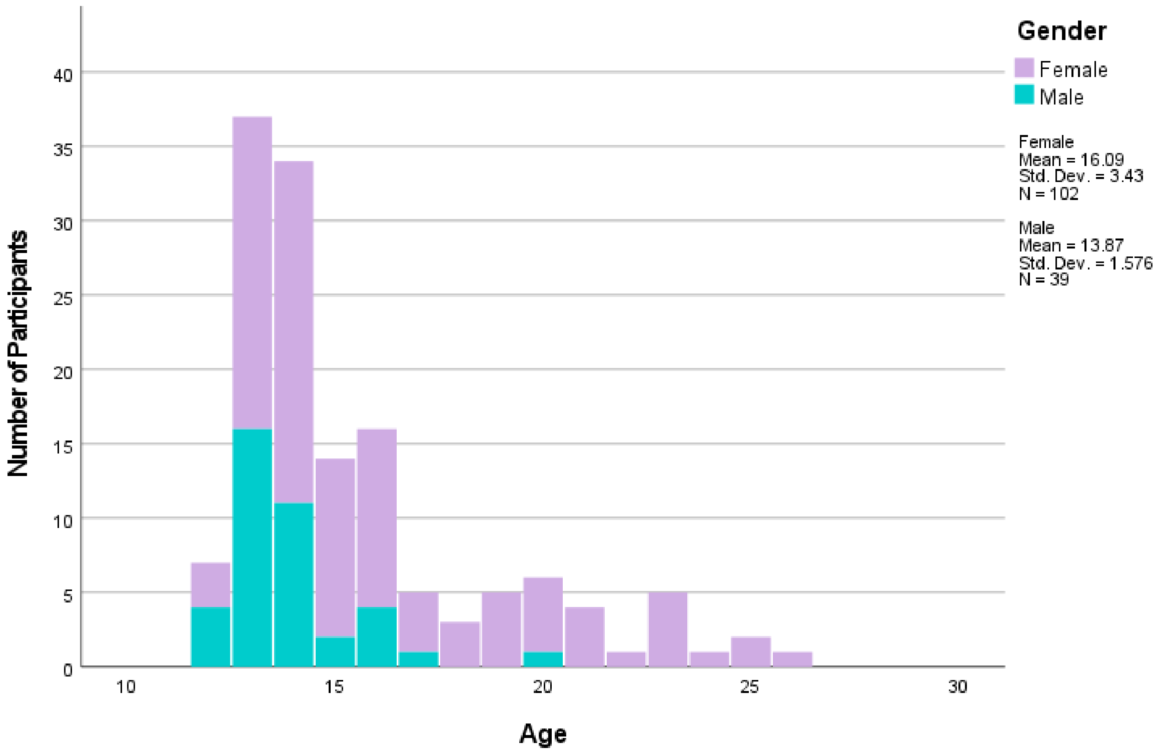
25. Analysis scripts (Optional)
    25.1. (Optional) Upload an analysis script with clear comments. This optional step is helpful in order to create a process that is completely transparent and increase the likelihood that your analysis can be replicated. We recommend that you run the code on a simulated dataset in order to check that it will run without errors.

Other

26. Other
    26.1. If there is any additional information that you feel needs to be included in your preregistration, please enter it here.

Appendix B

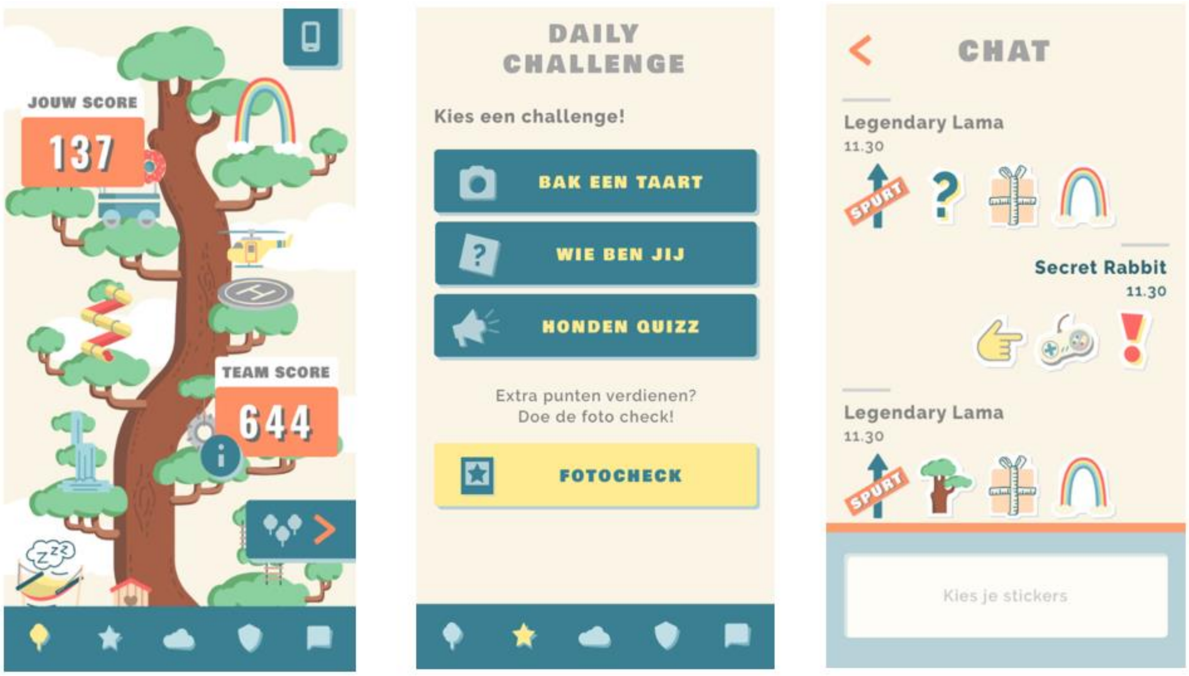
Distribution of Age and Gender of Cohort Three



Note. The histogram shows that the cohort ($N = 143$) is on the younger side. Most older participants of this cohort are female.

Appendix C

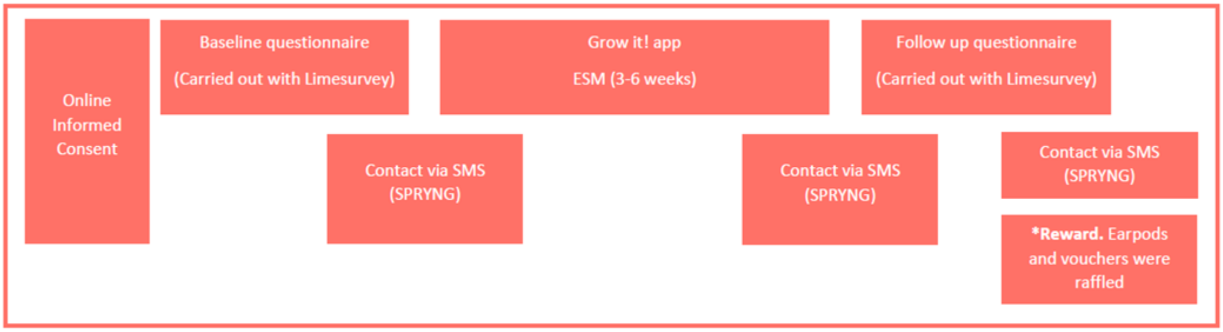
Screen Captures of the Grow It! App



Note. The left illustration shows the tree that users aim to grow. The little pictures in the tree are visual rewards. In the top left adolescents can see their own score, and the in the bottom right they can see their team score. The middle illustration shows the option menu for the daily challenges. Users can choose one out of three challenges: bake a cake, “who are you” and a dog quiz. Below is a “photo check” challenge to earn extra points. The right illustration shows the chat function. Teammates can communicate with stickers. Names are anonymized.

Appendix D

The Study Timeline



Note. Retrieved from the online codebook (*Grow It! Corona Project*, 2021).

Appendix E
Syntax

Hypothesis 1: The Paired Samples T-tests

Assumptions: Normality and Outliers

```
COMPUTE DifferenceAFF=WE01_2 - WE01_1.
EXECUTE.

COMPUTE DifferenceCOG=WE02_2 - WE02_1.
EXECUTE.

EXAMINE VARIABLES=DifferenceAFF
/PLOT BOXPLOT HISTOGRAM NPPLOT
/COMPARE GROUPS
/STATISTICS DESCRIPTIVES EXTREME
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.

EXAMINE VARIABLES=DifferenceCOG
/PLOT BOXPLOT HISTOGRAM NPPLOT
/COMPARE GROUPS
/STATISTICS DESCRIPTIVES EXTREME
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
```

Paired Samples T-tests

```
T-TEST PAIRS=WE01_2 WITH WE01_1 (PAIRED)
/ES DISPLAY(TRUE) STANDARDIZER(SD)
/CRITERIA=CI(.97500)
/MISSING=ANALYSIS.

T-TEST PAIRS=WE02_2 WITH WE02_1 (PAIRED)
/ES DISPLAY(TRUE) STANDARDIZER(SD)
```

```
/CRITERIA=CI(.97500)
/MISSING=ANALYSIS.
```

```
GRAPH
 /BAR(SIMPLE)=MEAN(DIFFAFF) MEAN(DIFFCOG)
/MISSING=LISTWISE
/INTERVAL CI(97.5)
/TITLE='Changes in Affective and Cognitive Well-being'.
```

Individual Effect Sizes

```
COMPUTE SDCOG=SD(WE02_1,WE02_2).
EXECUTE.
```

```
COMPUTE SDAFF=SD(WE01_1,WE01_2).
EXECUTE.
```

```
FREQUENCIES VARIABLES=SDAFF SDCOG DifferenceAFF DifferenceCOG
/ORDER=ANALYSIS.
```

Hypothesis 2: The One-way ANOVA

Assumption: Normality

```
EXAMINE VARIABLES=ESMp
/PLOT BOXPLOT HISTOGRAM NPPLOT
/COMPARE GROUPS
/STATISTICS DESCRIPTIVES EXTREME
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
```

Demographic Information and User Evaluation

Descriptives

```
DESCRIPTIVES VARIABLES=LEEFTIJD WE01_1 WE01_2 WE02_1 WE02_2 ESMp
esmresponses CIJFER_DESIGN CIJFER_GROWIT Hoeduidelijkvondjehetemotieoverzicht
/STATISTICS=MEAN STDDEV MIN MAX.
```

Frequencies

```
FREQUENCIES VARIABLES=GESLACHT ETH LEEFTIJD Welkonderwijsvolgjenu
Indienmiddelbareschoolwelkleerjaar Indienmiddelbareschoolwelkniveau Indienvervolgstudiewelkniveau
Ikleerdemezelfbeterkennen Ikvoeldemeerbeterdoor Ikdachtvakernaoverhoeikmevoel geeneffect anders
Hebjehetemotieoverzichtgedeeldmetanderen Watvooreffecthadhetemotieoverzichtopjemeerdereantwoo
Hoeduidelijkvondjehetemotieoverzicht AANBEVELEN Metwiehebjehetemotieoverzichtgedeeld
/ORDER=ANALYSIS.
```

Graph Age & Gender

```
GGRAPH
/GRAPHDATASET NAME="graphdataset" VARIABLES=LEEFTIJD GESLACHT
MISSING=LISTWISE REPORTMISSING=NO
/GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
SOURCE: s=userSource(id("graphdataset"))
DATA: LEEFTIJD=col(source(s), name("LEEFTIJD"))
DATA: GESLACHT=col(source(s), name("GESLACHT"), unit.category())
GUIDE: axis(dim(1), label("Age"))
GUIDE: axis(dim(2), label("Frequency"))
GUIDE: legend(aesthetic(aesthetic.color.interior), label("Gender"))
GUIDE: text.title(label("Stacked Bar of Age by Gender"))
SCALE: cat(aesthetic(aesthetic.color.interior), include("1", "2", "3"))
ELEMENT: interval.stack(position(summary.count(bin.rect(LEEFTIJD))), color.interior(GESLACHT),
shape.interior(shape.square))
END GPL.
```

Exploratory Correlations

```
CORRELATIONS
/VARIABLES = GESLACHT LEEFTIJD ESMp CIJFER_DESIGN CIJFER_GROWIT
Hoeduidelijkvondjehetemotieoverzicht DifferenceCOG DifferenceAFF
/PRINT = TWOTAIL.
/MISSING = LISTWISE.
```

Appendix F

Calculations ANOVA and Effect Size

ANOVA Calculation

I calculated the ANOVA online (https://statpages.info/anova1sm.html), which produced the following output.

Group Name	N (count)	Mean	Std. Dev.
Group 1	462	14.20	20.27
Group 2	733	20.56	25.79
Group 3	143	37.63	23.69
Group 4			
Group 5			
Group 6			
Group 7			
Group 8			
Group 9			
Group 10			

Desired confidence level for post-hoc confidence intervals: 95

Compute

ANOVA Table...

Source of Variation	Sum of Squares	d.f.	Variance	F	p
Between Groups:	60170.9149	2	30085.4575	53.1288	0.0000
Within Groups:	755975.9343	1335	566.2741		
Total:	816146.8492	1337			

Post-hoc tests...

```
Tukey HSD Post-hoc Test...
Group 1 vs Group 2: Diff=6.3600, 95%CI=3.0444 to 9.6756, p=0.0000
Group 1 vs Group 3: Diff=23.4300, 95%CI=18.0888 to 28.7712, p=0.0000
Group 2 vs Group 3: Diff=17.0700, 95%CI=11.9675 to 22.1725, p=0.0000
```

Effect Size Calculation of Grow It! Original Versus Grow It! Extended

To calculate an effect size that was relevant to the second hypothesis, I pooled cohort one and two were into one group: the group who used Grow It! original. I compared this group to cohort three, the group using Grow It! extended.

Pooling Cohort One and Two into One Group

Calculation Pooled Means. Pooled means of cohort one and two = (Mean1∗SampleSize1+Mean2∗SampleSize2)/(SampleSize1+SampleSize2)

= 14.20 * 462 + 20.56 * 733 / 462 + 733

= 6560.4 + 15070.48 / 1195

= 18.1

$M_{pooled(1+2)}$ = 18.1

Calculation Pooled Standard Deviations. Pooled standard deviations cohort one and two =

$$SD_{pooled} = \sqrt{\frac{(SD_1^2 + SD_2^2)}{2}}$$

= √ ((20.272 + 25.792)/ 2)

= √ (537.99)

= 23.19

$SD_{\text{pooled}(1+2)}$ = 23.19

Calculation of Cohen's d: Cohort One and Two Versus Cohort Three

The calculation required the following formula: Cohen's $d = (M_2 – M_1) / SD_{pooled}$.

Calculation Pooled Standard Deviations. $SD_{pooled\ (1\&2\ +\ 3)}$ =

= √ ((23.192 + 23.692)/2)

= √ (549.496)

= 23.44

$SD_{pooled\ (1\&2\ +\ 3)}$ = 23.44

Calculating Cohen's *d*. Cohen's $d = (M_3 – M_{pooled\ (1\&2)}) / SD_{pooled\ (1\&2\ +\ 3)}$

= (37.63 – 18.1) / 23.44

= 0.83

Cohen's *d* = 0.80, demonstrating a large effect**.**

Appendix G

Assumptions

Assumptions Hypothesis 1

Independence

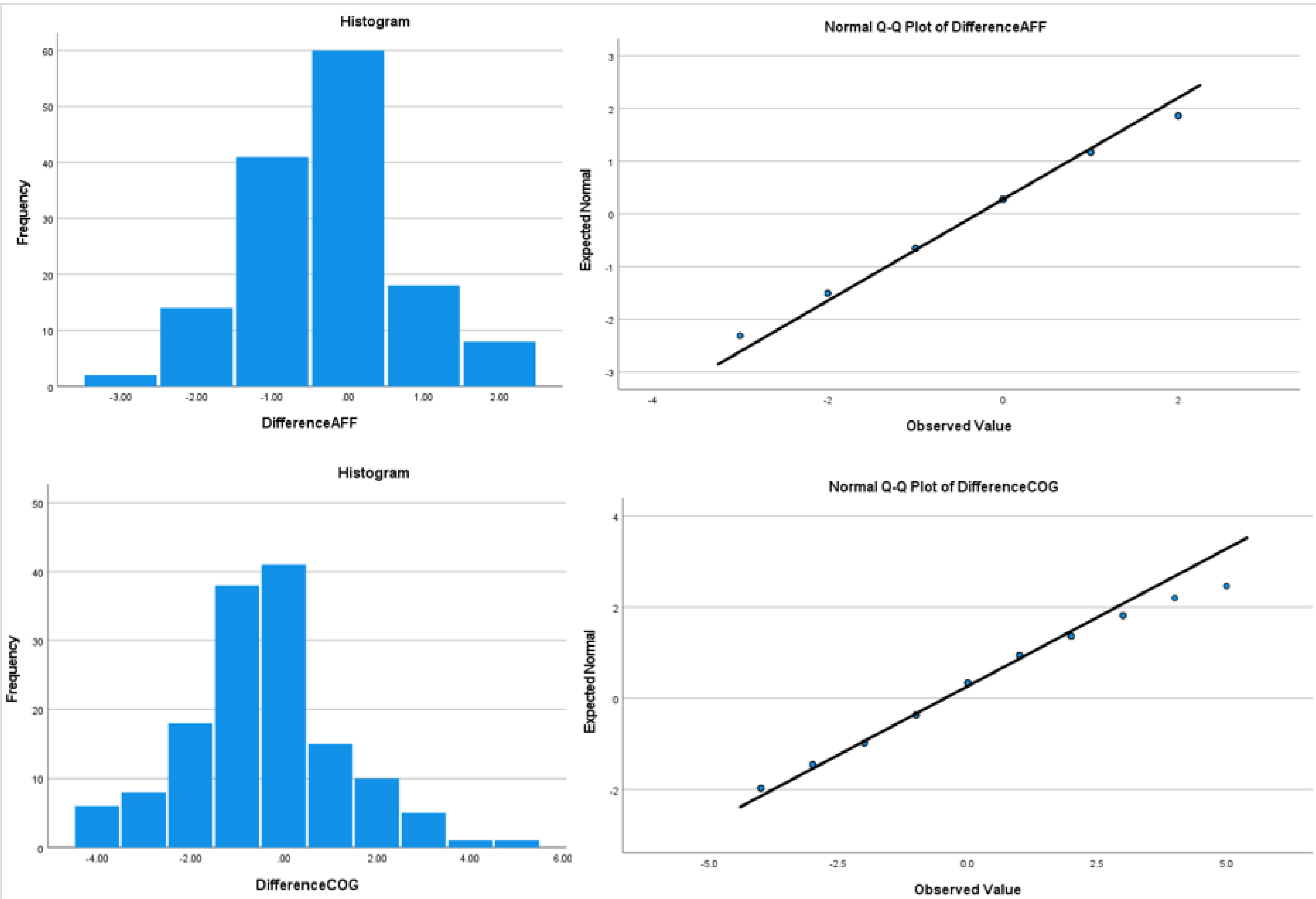
Each of the paired measurements were obtained from the same subject and subjects were unrelated amongst each other. Therefore, both cognitive and affective well-being met the assumption of independence.

Normality

The histograms (see Figure G1 below) look normally distributed as they exhibit a bell-shaped curve. QQ-plots show datapoints that are all relatively close to the line, indicating normality. Cognitive well-being differences indicate a small positive skew in datapoints. Absolute Skewness did not exceed .80 and absolute value of kurtosis did not exceed 2 either for affective (skewness = .07, kurtosis = .14) nor cognitive (skewness = .26, kurtosis = .62) well-being, indicating normality. The small positive skew in cognitive well-being differences was therefore considered negligible and likely not a problem considering the sample size. Furthermore, formal analyses demonstrated an absence of normality. Both the Shapiro-Wilk as well as the Kolmogorov-Smirnov tests of normality rejected the null hypothesis of normality ($p < .001$) for both affective and cognitive well-being. I did not take formal tests into account as they tend to be sensitive to slight deviations in normality and because paired samples t-tests are robust against this. Taken together, the results confirmed the assumption of normality.

Figure G1

Normality in Affective and Cognitive Well-being Differences



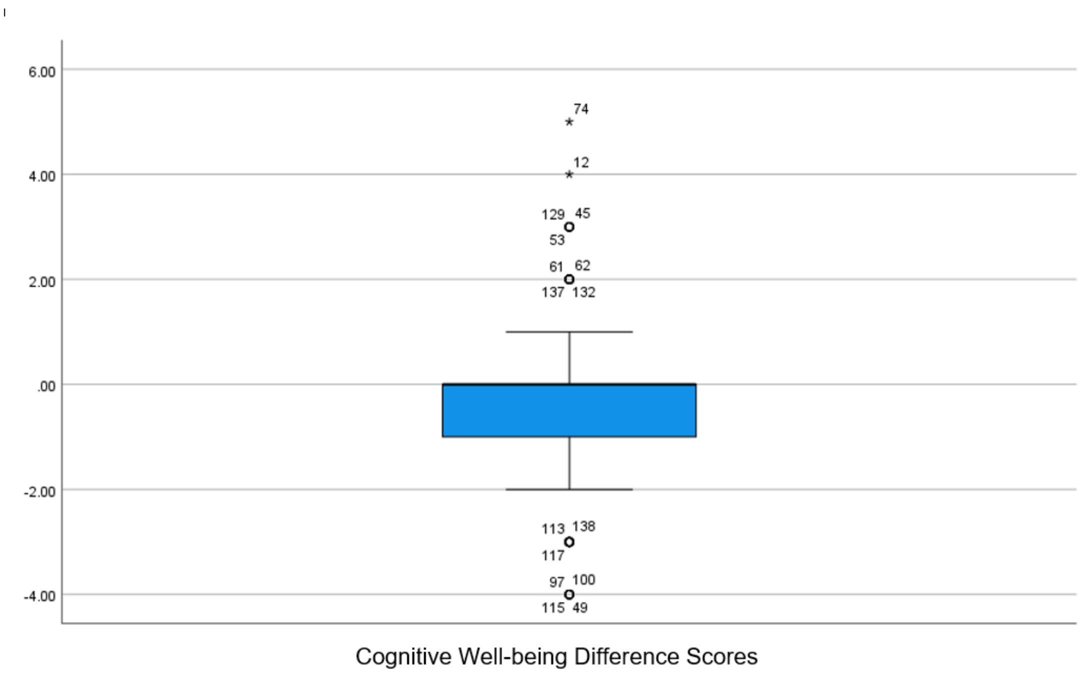
Note. The illustration above shows the histograms and QQ-plots that I inspected for normality. Inspections for affective well-being scores, labelled as ‘DifferenceAFF’, are represented in the upper left and right. For cognitive well-being scores, labelled as ‘DifferenceCOG”, I inspected the lower left and right graphs. All graphs illustrate relatively normal distributions.

Extreme Outliers

There were no extreme outliers for affective well-being differences. There were two extreme outliers for cognitive well-being differences, violating the assumptions of no extreme outliers (see Figure G2). After inspecting the cases I determined there were no data-entry errors. Conducting the analysis with the extreme outliers would likely not impact the t-test results. This idea is supported by Van Selst and Jolicoeur (1994), as they demonstrate that in larger sample sizes outliers have relatively smaller impact on dispersion and central tendency. Therefore, I did not remove the outliers.

Figure G2

Extreme Outliers in Cognitive Well-being Differences



Note. Illustrated above is a boxplot of cognitive well-being difference scores. Multiple outliers are present, with two outliers extremely deviating from the normal distribution.

Assumptions Hypothesis 2

Independence

A small chance existed that subjects were not entirely independent from each other as they may have originated from the same school, classroom, or family. Nevertheless, nothing specifically indicated that they were not independent in each cohort. Therefore, the assumption of independence was met.

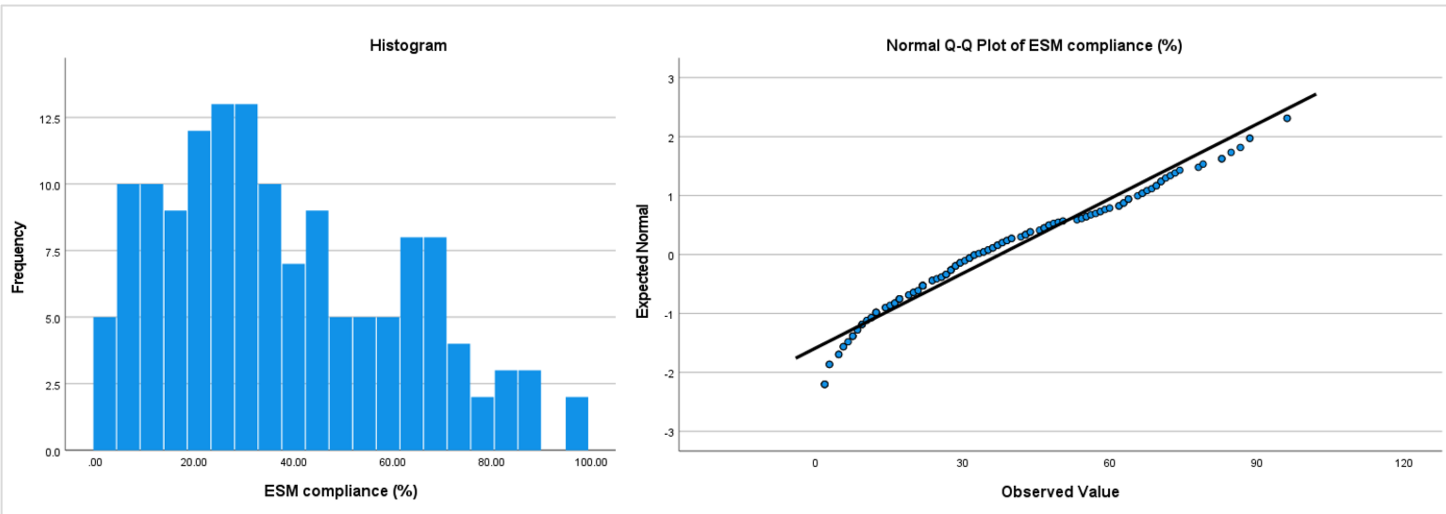
Normality

Regarding ESM compliance percentages, the histograms look slightly positively skewed, indicating non-normality. QQ- plot shows datapoints are all relatively close to the line indicating normality (see Figure G3). The absolute Skewness did not exceed .80 and the absolute value of kurtosis did not exceed 2 either (skewness = .52, kurtosis = .60). Therefore, the assumption of normality was met. The Shapiro-Wilk (*p* =

.004) as well as the Kolmogorov-Smirnov ($p < .001$) tests of normality rejected the null hypothesis of normality. Nevertheless, formal tests were not considered due to sensitivity in detecting deviations in normality as well as the ANOVA being robust against this.

Figure G3

Normality in ESM Compliance



Note. The illustration above shows the histogram and QQ-plot that I inspected for normality in the variable ESM compliance. The distribution shows a small positive skew.

Homogeneity of Variances

This assumption was untestable as I received no access to the data of cohort one and two. Only the summary data (mean, standard deviation, and sample size) was available, which was not sufficient to manually calculate the Levene's test. The assumption of homogeneity of variances was not met.