MASTERARBEIT | MASTER'S THESIS

Titel | Title

Customizing Success: Investigating the Impact of Tracking Schedules and User Characteristics on Digital Food Monitoring Adherence

verfasst von | submitted by

Maja Ruby Gabriel BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Psychologie
Betreut von | Supervisor:	Univ.-Prof. Dr. Laura Maria König B.Sc. M.Sc.

Abstract

Fettleibigkeit und schlechte Ernährung sind nach wie vor große globale Gesundheitsprobleme. Obwohl sich digitales Ernährungs-Tracking (*digital food tracking*) als wirksame Maßnahme zur Verbesserung des Ernährungsbewusstseins und des Gewichtsmanagements gezeigt hat, fällt Menschen das langfristige Tracking der eigenen Ernährung schwer. In dieser Studie werden Faktoren untersucht, die die Adhärenz und Wirksamkeit der digitalen Ernährungsüberwachung durch *Self-Monitoring* in einem *Within-Between-Subjects-Design* beeinflussen. Die Daten wurden mittels eines *Ecological Momentary Assessments* erhoben. Über einen Zeitraum von 30 Tagen erfassten 167 Teilnehmer*innen ihren Obst- und Gemüsekonsum über eine Smartphone-App gemäß drei verschiedenen Tracking Plänen (selbst gewählte Tracking-Tage, Tracking nur an Wochentagen oder zufällig zugewiesene Tracking-Tage). Anhand von Daten aus Fragebögen, welche zu Beginn und am Ende der Studie ausgefüllt wurden, wurde der Einfluss von Gewissenhaftigkeit und wahrgenommener Autonomie auf die Adhärenz und die Wirksamkeit der Studie (gemessen als Veränderung des Obst- und Gemüsekonsums im Verlauf der Studie) untersucht. Die Ergebnisse zeigten keine Unterschiede in der Einhaltung der Tracking Pläne, wobei die Adhärenz in allen Gruppen hoch war und kein verringertes Tracking an Wochenenden festgestellt wurde.

Hinsichtlich wahrgenommener Autonomie zeigten die Ergebnisse keine Unterschiede zwischen den drei Versuchsgruppen. In allen Versuchsgruppen lagen hohe Ausprägungen in wahrgenommener Autonomie vor. Auch hinsichtlich der Adhärenz unterschieden sich die Versuchsgruppen nicht, wobei die Adhärenz in allen Gruppen hoch ausgeprägt war und kein erkennbarer Wochenend-Effekt sichtbar wurde. Wie erwartet, stieg der Obst- und Gemüsekonsum im Laufe der Studie signifikant an. Jedoch konnte kein Zusammenhang zwischen Adhärenz und Studienwirksamkeit gefunden werden. Auch Gewissenhaftigkeit hatte keinen Einfluss auf die Adhärenz oder den Obst- und Gemüsekonsum.

Es muss jedoch erwähnt werden, dass die statistische Variabilität in allen relevanten Variablen begrenzt war, was statistisch die Fähigkeit zur Erkennung von Zusammenhängen beeinträchtigt. Daher ist unklar, ob das Fehlen signifikanter Zusammenhänge auf einen tatsächlichen Mangel an Zusammenhängen oder auf eine unzureichende Variabilität der Daten zurückzuführen ist. Diese Ergebnisse unterstreichen die Komplexität der Faktoren, die die digitales *Self-Monitoring* beeinflussen, und legen nahe, dass zukünftige Forschung diese Effekte in repräsentativeren Stichproben untersuchen sollten, um die vorliegenden Ergebnisse besser einordnen zu können, da diese teilweise im Widerspruch zur bisherigen Literatur stehen.

Abstract

Obesity and poor nutrition remain major global health concerns. Although digital food tracking has proven effective for nutritional awareness and weight management, maintaining adherence is challenging. This study examined factors influencing adherence and efficacy of digital food tracking using self-monitoring in a within-between-subjects design. Data was obtained through an Ecological Momentary Assessment. Over 30 days, 167 participants tracked their fruit and vegetable intake on a smartphone app according to three tracking schedules (self-selected, weekdays only, or randomly assigned). Pre- and post-questionnaire data were used to furthermore assess the influence of conscientiousness and perceived autonomy on adherence and study efficacy (measured as a change in fruit and vegetable consumption over the course of the study).

Results indicated no differences in perceived autonomy, with high levels of perceived autonomy across all groups. Likewise adherence rates were equal across all tracking schedules, with high adherence across all groups and no evident *weekend-effect*. As expected, fruit and vegetable consumption did significantly increase over the course of the study. However adherence did not predict study efficacy and conscientiousness did not moderate adherence or fruit and vegetable consumption.

Across all relevant variables, statistical variability was limited, which likely reduced the ability to detect associations. Consequently, it remains unclear whether the absence of significant relationships reflects a genuine lack of association or insufficient variability in the data. These findings highlight the complex nature of factors influencing digital self-monitoring and suggest that future research should examine these effects in more diverse and representative samples to clarify the present results, which partially contradict prior literature.

Index

Introduction

According to the 2019 Health Survey by Statistics Austria, only 3.8% of men and 7.1% of women meet the recommended daily intake of fruits and vegetables (Statistik Austria, 2025). This issue is not limited to Austria. On a global scale, obesity and poor nutrition are among the biggest health challenges due to their strong association with secondary diseases such as type 2 diabetes, cardiovascular disease, stroke, and certain types of cancer (Rush & Yan, 2017). In response, considerable research has focused on developing effective strategies to promote healthier dietary behaviors. App-based mobile health interventions have emerged as a promising approach to improving nutritional behaviors and related health outcomes, largely due to the widespread global use of smartphones, the ease of app installation across devices, and their typically free availability, which enhances healthcare accessibility for a broad population (Bardus et al., 2016; Villinger et al., 2019).

To date, most users have depended on app-based mobile health interventions, which have undergone little empirical evaluation and rarely incorporate evidence-based behavior change strategies (Rivera et al., 2016). Behavior change strategies are techniques used to support individuals in adopting and maintaining healthy behaviors, including healthier eating behavior (Hongu et al., 2011).

Among health interventions that incorporate behavior change strategies, self-monitoring is not only one of the most frequently utilized behavior change techniques (BCT) in mobile app-based dietary interventions but also among the most effective BCTs (Berry et al., 2021; Burke et al., 2011; Michie et al., 2009; Villinger et al., 2019). Self-monitoring refers to the process where individuals track their own health-related behaviors and adherence to doctors orders or health behavior guidelines, empowering them to manage their own health behavior (Kgasi et al., 2024).

In the meta-analysis by Patel et al. (2021), 74% of the included studies reported a connection between frequent self-monitoring and weight loss. Furthermore digital self-monitoring, particularly food tracking, has shown to be highly effective in enhancing dietary awareness and supporting weight management (Patel et al., 2021). Additionally, self-monitoring facilitates behavior change in eating habits, including increased fruit and vegetable consumption (Gardiner & Bryan, 2021).

Although self-monitoring is linked to positive outcomes, achieving consistent adherence to tracking remains a challenge for individuals (Harvey et al., 2019; Turner-McGrievy et al., 2019). While the World Health Organization (2003) defines adherence as the extent to which an individual's behavior aligns with healthcare professional recommendations, dietary tracking research often operationalizes adherence through metrics such as the frequency of logged entries, the number of recorded food items, or the number of documented meals (Harvey et al., 2019; Lee et al., 2023; Pellegrini et al., 2018).

König et al. (2021) identified several barriers that contribute to the inconsistent use of nutrition apps, highlighting the need for solutions that promote long term user engagement. These barriers encompass user-specific factors like goal-setting or motivation as well as app-specific factors which include app usability, food database and technical issues (König et al., 2021).

The Impact of Adherence on Goal Achievement

Goal achievement is the result of a continuous process of self-regulation (Inzlicht et al., 2021). Self-regulation refers to the dynamic process of determining a desired end state (i.e., a goal) and then taking action to move toward it while monitoring progress along the way (Inzlicht et al., 2021).

The cybernetic model of self-regulation by Carver & Scheier (1998) is one of the most influential self-regulation theories and breaks down this abstract concept into four key elements which are interconnected by a feedback loop. The first key element is the defined goal. The second element is input about the current state of the organism regarding that goal. The third element is a system that monitors for discrepancies between the goal and the current state. While the last element is a system that implements changes to reduce the discrepancies between the goal and the current state.

In literature the term self-monitoring is often based on the third element of the cybernetic model of self-regulation by Carver & Scheier (1998). As previously mentioned, particularly frequent self-monitoring is linked to goal achievement (Patel et al.,2021). In the context of the cybernetic model, adherence to self-monitoring leads to a more frequent feedback loop. Therefore adherence does not merely reflect behavioral consistency, it actively facilitates goal achievement by sustaining the feedback processes described in the cybernetic model. Consequently, adherence can be understood as both an indicator and a mechanism of successful self-regulation. This theoretical framework may explain why self-monitoring interventions are frequently associated with positive health outcomes, such as improved dietary behaviors (Mossavar-Rahmani et al. 2004).

The Impact of Autonomy on Tracking Adherence

Based on their systematic review, König et al. (2021) define design guidelines that are intended to promote adherence with food tracking apps. One of these recommendations is to make the app customisable in order to take individual needs and goals into account. While the recommendation is based on empirical findings, it can also be theoretically supported by Ryan and Deci's (2000) Self-Determination Theory (SDT).
The self-determination theory (SDT) by Ryan and Deci (2000) states that intrinsic motivation is driven by the fulfillment of the need for autonomy, competence and relatedness. According to the SDT, these three fundamental needs serve as primary drivers of intrinsic motivation and significantly influence human behavior.
In the health context, greater autonomy has been linked to improved treatment adherence and motivation (Flaherty et al., 2019; Glusman et al., 2019; Kennedy et al., 2004). Behavior change is more likely when people are intrinsically motivated, which means that they are driven by personal interest and internal values, rather than by external pressures (Ng et al., 2012). Based on these findings, Flaherty et al. (2019) suggest that food tracking apps may enhance user adherence by allowing personalization. Customizable features can promote perceived autonomy and, in turn, foster intrinsic motivation and continued app engagement.

The Impact of Weekdays on Tracking Adherence

Studies examining adherence indicate that there are systematic differences in adherence on certain days of the week. Regardless of the specific activity being scheduled, adherence tends to be lower on weekends compared to weekdays. For instance, adherence to medication intake schedules is often reduced on weekends (Vervloet et al., 2013), as is adherence to physical activity recommendations and sports schedules (Gubelmann et al., 2017; Galindo-Perdomo et al., 2021).

Similarly, in the area of food tracking, research findings indicate that weekends present challenges in maintaining adherence to tracking schedules (Krukowski et al., 2013; Pellegrini et al., 2018). Pellegrini and colleagues (2018) referred to this phenomenon as *weekend effect*. Since this phenomenon occurs regardless of the specific activity being scheduled and appears to be closely linked to adherence, in this study one of the trial groups is designed to eliminate the so-called weekend effect by excluding weekends from the tracking schedule.

The Impact of Conscientiousness on Tracking Adherence and Goal Achievement

Conscientiousness is a personality trait characterized by self-discipline, orderliness, goal orientation, and a strong sense of responsibility. Research shows that this personality trait is strongly correlated with higher adherence to health-related behaviors (Soto et al., 2015). Individuals with higher levels of conscientiousness are more likely to follow medical recommendations and adopt health behaviors such as regular exercise and healthy eating, which contribute to physical well-being and increased longevity (Bogg & Roberts, 2004; Gartland et al., 2020). Furthermore conscientious individuals tend to show higher levels of adherence to exercise schedules (Lakshmanan et al., 2024) as well as greater adherence to dietary guidelines, including increased consumption of fruits and vegetables and reduced alcohol intake (Gartland et al., 2020; Kern & Friedman, 2008). However, while these associations are well documented in broader health behavior contexts, conscientiousness has rarely been examined as a moderating factor in studies on digital self-monitoring. This is particularly evident in the context of food tracking.

Furthermore conscientiousness plays a crucial role in the attainment and maintenance of health-related goals. Research shows that individuals high in conscientiousness are more likely to sustain health behaviors over time (Chopik & Lee, 2022) and to regulate lifestyle habits more effectively (Gold et al., 2020). Notably, those who maintain weight loss over the long term tend to exhibit higher levels of conscientiousness than individuals who regain weight, underscoring the trait's importance in sustaining behavioral change (Gold et al., 2020). Similar to the association between conscientiousness and adherence, the link between conscientiousness and goal achievement has been explored primarily in broader health contexts, yet both associations remain underexamined within digital self-monitoring frameworks. Given the consistent findings of the role of conscientiousness in the context of health behaviors, investigating conscientiousness as a potential moderator may provide valuable insights into which individuals benefit most from digital self-monitoring.

Present study

Research questions and hypotheses

The objective of the study is to assess how digital food tracking can be facilitated by investigating user-related as well as app-related factors that may influence adherence and efficacy of food-tracking applications. Based on the research findings presented above, five research questions and their corresponding hypotheses were derived. As a methodological basis serve three different self-monitoring schedules for food tracking, which were operationalized as the three experimental groups. Group 1 was designed to enhance the feeling of autonomy by allowing participants to choose their own tracking days. To account for the *weekend effect* described by Pellegrini et al. (2018) participants of Group 2 tracked exclusively on weekdays (Monday to Friday). Group 3 served as the control group, following tracking schedules with randomly selected days.

The following section presents the research questions and the corresponding hypotheses. The first research question evaluates how the different self-monitoring schedules influence perceived autonomy over the course of the program. It is hypothesized that participants in Group 1, who chose their own tracking days, would report higher levels of perceived autonomy than those in Group 2 and Group 3 (Hypothesis 1).

The second research question investigates whether different self-monitoring schedules result in varying adherence rates. Based on the autonomy findings, participants in Group 1, who chose their own tracking days, are expected to show higher adherence than participants in Group 3 whose tracking schedules were randomly assigned (Hypothesis 2a). Considering the weekend effect, participants in Group 2 who tracked exclusively on weekdays are expected to show higher adherence than participants in Group 3 (randomly assigned tracking days; Hypothesis 2b). Additionally, participants in Group 3 are expected to demonstrate higher adherence on weekdays than on weekends (Hypothesis 2c).

The third research question investigates whether the different self-monitoring schedules lead to an increased fruit and vegetable intake (serving as a measure of the study's efficacy). The corresponding hypothesis is that all three self-monitoring plans lead to an increase in fruit and vegetable consumption (Hypothesis 3).

The fourth research question examines the relationship between study efficacy and adherence across the different trial groups. Based on the previous research the assumption is that there is a positive relationship between study efficacy (fruit and vegetable consumption) and adherence regardless of the trial group (Hypothesis 4).

The fifth and last research question examines whether conscientiousness moderates the effects of the various tracking schedules on both study efficacy (fruit and vegetable consumption) and adherence. Specifically, it is hypothesized that conscientiousness moderates the impact of the different food-tracking schedules on adherence, predicting that participants with higher conscientiousness exhibit greater adherence (Hypothesis 5a). In addition, it is assumed that conscientiousness moderates the impact of the food-tracking plans on study efficacy, with more conscientious participants showing higher study efficacy (Hypothesis 5b).

Methods

This study is part of a larger research project under the supervision of Univ.-Prof. Dr. Laura M. König which was approved by the ethics committee of the University of Vienna and was preregistered in February 2024 using the “Preregistration for Quantitative Research in Psychology (PRP-QUANT) Template” prior to creation of data.

Furthermore this study builds on the pilot study “Burden of Food Tracking: Do Different Schedules for Self-Monitoring Have an Impact on the Adherence to Digital Food Tracking and Why?” conducted by Maja Gabriel, Maja Halmburger, Tanja König, Rosa List, Paula Moosbauer, Karoline Moser, and Moritz Tiemann. The pilot study included only 53 participants and primarily addressed the first two research questions of the current study. Due to the limited sample size, effect sizes could not be estimated, which highlighted the need for a follow-up study with a larger sample to examine whether the results hold under more robust conditions.

The current study includes a few deviations from the preregistration, which are documented and explained in Table 1. Furthermore they are addressed in the respective sections of the paper.

Participants

An a priori power analysis was conducted in G*Power 3.1.9.7 to determine the sample size required for 80% power. Because comparable studies did not report effect sizes, the recommendation of Clark-Carter (2007) to assume a medium effect size (d = .50) based on Cohen’s guidelines served as a calculation basis. The between-subjects component of the mixed ANOVA for Hypothesis 2 demanded the largest sample of all planned analyses and thus served as the basis for the calculation, which indicated that N = 159 participants were needed (See Appendix K). Data collection concluded once 159 participants completed the post-questionnaire.

Participant recruitment (February 2024 - April 2025) took place through social media platforms (snowball sampling) as well as the Laboratory Administration System of Behavioral Science (LABS), yielding a convenience sample (see Appendix A for recruitment materials). LABS is a digital platform developed and operated by the University of Vienna to facilitate the mandatory participation of psychology students in psychological studies of the university. Each psychology student at the University of Vienna is required to earn LABS credits during their Bachelor’s degree program. Participants recruited through LABS received credits as compensation for their participation, while all other participants did not receive any compensation.

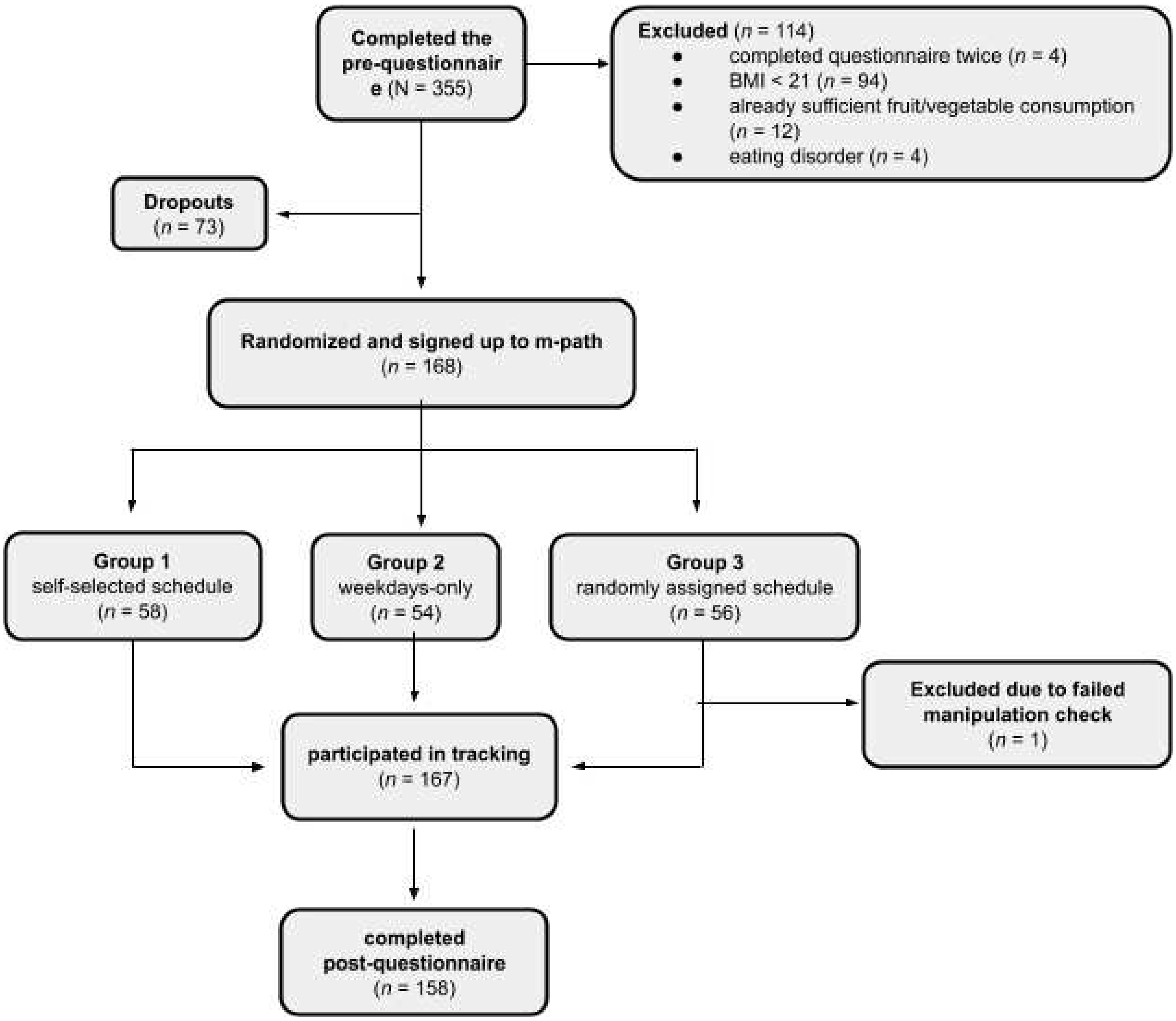
The initial eligibility of participants was evaluated using an online questionnaire. Informed consent was obtained both via the questionnaire and during the subsequent online meeting. Additionally, participants submitted a signed written consent form. Specific inclusion and exclusion criteria were established to prioritize participant safety and ensure their relevance to the research objectives. Eligibility requirements included a body mass index (BMI) of at least 21, calculated as weight (kg) divided by height squared (m^2). This study focused on monitoring fruit and vegetable portions through food tracking, and therefore, no weight loss was expected. By establishing a minimum BMI of 21, the aim was to further reduce the risk of exceeding the healthy BMI threshold (Kaiser et al., 2014).

Furthermore participants needed to be at least 18 years old, fluent in German, and in possession of a smartphone with an operating system compatible with the application (Android or Apple). Motivation to increase fruit and vegetable intake was ensured by excluding individuals who already consumed more than two portions of fruit and three portions of vegetables per day, since they already meet the current diet recommendation (Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz, 2024). Exclusion criteria also included a current or past history of eating disorders, participation in weight management programs, the use of other dietary applications, or involvement in other studies related to eating behavior during the course of this study.

Of the 355 participants who completed the pre-questionnaire, 241 met all predefined inclusion criteria. A total of 73 individuals could not be reached for study onboarding and tracking app installation, and were therefore classified as dropouts. Consequently, 168 participants were successfully enrolled into the tracking app. Participants were assigned to the three experimental groups using block randomization (blocks of nine) leading to a distribution of n = 58 in Group 1 (self-selected schedule), n = 54 in Group 2 (weekdays only), and n = 56 in Group 3 (randomly assigned). One participant in Group 3 was excluded from the analysis due to consistently tracking on incorrect days, resulting in an adherence score of zero. Therefore this individual did not pass the instruction manipulation check, which led to a reduction in the final sample size from 168 to 167. Of these 167 participants, 158 completed the post-questionnaire. A detailed overview of the exclusion process is provided in Figure 1.

Figure 1.

Flow diagram of participant selection and drop outs



Study Design and Procedure

This experimental study employed a within-between-subject design where the within-subject factor was the change in fruit and vegetable consumption over time of each participant. The between-subjects factor was the comparison between the three trial groups.

People who were interested in the study initially completed an online questionnaire (see Appendix B). This pre-questionnaire functions both as a screening instrument to evaluate participants' eligibility according to inclusion and exclusion criteria and as a source of information for the study's progression and subsequent analysis. People were informed about the purpose and process of the study, as well as how their data would be handled, including aspects of anonymity and data protection. Furthermore this questionnaire screened for exclusion criteria, gathered sociodemographic data, assessed participants' average fruit and vegetable consumption over the previous week, evaluated their attitudes toward eating behavior, and measured their levels of conscientiousness. If any participant's responses indicated that they met an exclusion criterion, they were given a brief explanation for their exclusion, along with our gratitude for their interest in the study. Participants who did not meet any exclusion criteria were required to create a passcode to ensure anonymity and to enable their pre-questionnaire data to be later linked with the data from the tracking app.

Participants who were not recruited through LABS provided their email addresses and were subsequently contacted via email to schedule the upcoming online meeting, while those recruited through LABS arranged their appointment through the LABS system. Before the emails were sent, participants were randomly assigned (using block randomization in blocks of nine) to one of the three self-monitoring groups using the randomization program *random.org* provided by Haahr (o. D.). In the email received by the participants, they were asked to sign and return the consent form. Group 1, which was allowed to create their own tracking schedule, was additionally prompted to identify 20 days within the next 30 days (starting from the date of the upcoming online meeting) on which they would be able to track their fruit and vegetable consumption (see Appendix C for the emails sent to participants and Appendix D for the consent form). Also prior to the meeting, the personalized tracking schedules for participants in groups 2 and 3 were prepared, while the tracking schedule for participants in Group 1 was created during the upcoming online meeting (see Appendix E for examples of tracking schedules). The tracking schedule for participants in Group 2 consisted of 20 consecutive days, starting the day after the upcoming online meeting, excluding weekends. Participants in Group 3 received a tracking schedule comprising 20 randomly allocated days within a 30-day timeframe, starting the day after the upcoming online meeting. This schedule was generated using the Python programming language (see Appendix H). This approach allowed participants to start the study at different times.

During the online meeting, participants were provided with detailed information on monitoring their fruit and vegetable intake and guidance on using the smartphone application. Furthermore, participants in Groups 2 and 3 were presented with their personalized tracking schedules, while the tracking schedule for participants in Group 1 was collaboratively created during the meeting (see Appendix E for examples of tracking schedules). Additionally, participants completed the welcome questionnaire within the app, where they were only required to enter their personal passcode. This passcode, previously provided in the pre-questionnaire, was used to link the data from the pre-questionnaire with the app data while ensuring complete anonymization.

Immediately following the meeting, participants received their tracking schedules along with an informational document on fruit and vegetable intake via email (see Appendix F for the emails sent and Appendix G for the document). Over a 30-day period, participants used the application to record their fruit and vegetable intake on 20 days. For each entry, they specified the type of meal (breakfast, lunch, dinner, or snack) and the portion sizes, ranging from 0 to 2 for fruit and 0 to 3 for vegetables, in steps of 0.5.

At the conclusion of the tracking period, participants completed a post-questionnaire integrated within the app (see Appendix J for Post-questionnaire). This questionnaire gathered data on participants' current weight, their average consumption of fruits and vegetables over the past seven days, their level of nutritional awareness, and the extent to which their need for autonomy was satisfied in the context of the tracking app. Additionally, the questionnaire included further items related to their overall experience with the tracking process. Upon completing the post-questionnaire, participants were then debriefed on the study’s aims and group assignment. While data collection was conducted anonymously, neither allocation concealment nor blinding of the researchers involved was applied.

Materials and Measures

Pre-questionnaire

Participants initially completed an online questionnaire via Unipark (Tivian XI GmbH, 2021). After providing informed consent, participants were asked to report descriptive data such as age and gender, as well as body mass index (BMI) calculated as weight (kg) divided by height squared (m^2). Additionally, they were asked to provide information about current or past history of eating disorders. This was followed by a self-assessment of conscientiousness which was essential for testing Hypotheses 5a and 5b.

Measurement of conscientiousness. Conscientiousness was operationalized using two items from the German short version of the Big Five Inventory-10 (BFI-10). Participants responded to the items “I see myself as someone who tends to be lazy” (German: “Ich bin bequem, neige zu Faulheit”) and “I see myself as someone who does a thorough job” (German: “Ich erledige Aufgaben gründlich”) on a 5-point Likert scale ranging from “disagree strongly” to “agree strongly.” To assess individual levels of conscientiousness, the responses to the two items were averaged. As the first item is negatively worded, it was reverse-coded prior to calculation. The resulting mean score represents a continuous measure of the personality trait conscientiousness, reflecting the extent to which individuals exhibit this trait rather than their agreement with specific statements. The categorization into low, medium, and high levels of conscientiousness was based on an equal division of the possible scale range (1-5), with each range spanning 1.33 points: 1.00 - 2.33 = low, 2.34 - 3.66 = medium, and 3.67 - 5.00 = high. As the scale consists of only two items, assessed at a single time point, the calculation of internal consistency was not appropriate, and test-retest reliability could not be evaluated in the present study. Instead, a correlation between the two items was computed, yielding a coefficient of r = .316, which is considered acceptable according to commonly used benchmarks: values around .10 indicate a small effect, around .30 a medium effect, and .50 or above a large effect (Ellis, 2010). Moreover, test-retest reliability for these items has already been established by Rammstedt et al. (2013).

Following the self-assessment of conscientiousness, participants received information about the current dietary recommendations of the Austrian Ministry of Health regarding fruit and vegetable consumption, along with guidance on how to estimate portion sizes (Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz, 2024). Participants then completed a self-assessment of their average daily fruit and vegetable intake over the past week which was later used for the computation of study efficacy.

Measurement of study efficacy. Study efficacy was operationalized as a significant increase in fruit and vegetable consumption over the course of the study. This outcome was essential for testing Hypotheses H3, H4, and H5b. To assess this, self-reported intake from one week prior to the intervention (pre-questionnaire) was compared with intake from the week preceding the post-questionnaire. Specifically, the total reported portions of fruit and vegetables from the pre-questionnaire were compared with the corresponding totals from the post-questionnaire. Participants indicated their average daily consumption using fixed portion categories. For fruit intake, the options were: "0," "1," "2," or "more than 2" portions per day. For vegetable intake, the options were: "0," "1," "2," "3," or "more than 3" portions per day. The difference in total intake between the two time points served as the indicator of change. A positive and statistically significant difference score was interpreted as evidence of study efficacy, reflecting an increase in fruit and vegetable consumption over the course of the intervention.

Study onboarding

Following the online meeting, where participants received detailed information about the study procedure and the food tracking application, they were sent one of three email templates based on their group allocation (see Appendix F). Each email included an information sheet outlining the recommended daily intake of fruits and vegetables, instructions on how to measure portion sizes, and an example of a meal plan that meets the guidelines (see Appendix G).

All tracking schedules were manually created by the study coordinators. The plans varied based on participants' group membership and the date of their online meeting, allowing participants to start the study at different times. Schedules for Group 1 were developed in consultation with the participants, while schedules for Group 2 consisted of twenty weekdays following the online meeting. The tracking schedules for participants in Group 3, who received a tracking plan with 20 randomly assigned days within the 30-day period, were generated using the Python programming language (see Appendix H).

Tracking-device

Food tracking was conducted using the "m-path" application, a free and customizable research tool developed by KU Leuven in the Netherlands (m-path, n.d.). For this study, a personalized set of surveys was created that participants could access daily for tracking their consumption. Participants had the option to track breakfast, lunch, dinner, and snacks. After selecting one of these categories, participants were first asked to report the vegetable content of their meal by choosing a portion size from the following options: “0”, “1”, “2”, “3” or “more than 3”. If they selected a portion size greater than 0, a text field appeared prompting them to specify the type of vegetable and its method of preparation. The same procedure was then applied to assess the fruit content of the meal, using the portion size options: “0”, “1”, “2” or “more than 2”. Screenshots of the app can be found in Appendix I.

Post-questionnaire

At the end of the 30-day period, participants were asked to complete a post-questionnaire within the app (see Appendix J for Post-questionnaire). This questionnaire assessed their weight, their self-reported average daily fruit and vegetable intake over the past week, and the extent to which their need for autonomy was satisfied in the context of the tracking app.
Their self-reported average daily fruit and vegetable intake over the past week was measured with the same item as in the pre-questionnaire. The degree of autonomy need fulfillment was assessed using an adapted version of the *Balanced Measure of Psychological Needs Scale* (Neubauer & Voss, 2016).

Measurement of perceived autonomy. Perceived autonomy was operationalized using an adapted version of the *Balanced Measure of Psychological Needs Scale* (Neubauer & Voss, 2016), specifically the subscale measuring autonomy need fulfillment, one of the core psychological needs represented in the scale. In this adaptation, participants were asked specifically about their perceived fulfillment of the need for autonomy in the context of using the app, rather than in general (Neubauer & Voss, 2016; Appendix J). This measure was essential for testing Hypothesis 1. The adaptation process is illustrated by the item “I was free to do things my own way” (German: “Ich hatte den Freiraum, Dinge so zu tun, wie ich es wollte”), which was modified for this study to “When I used the study app, I had the freedom to do things the way I wanted” (German: “Wenn ich die Studienapp benutzte, hatte ich den Freiraum, Dinge so zu tun, wie ich es wollte”).

The autonomy need fulfillment subscale assesses both the satisfaction and dissatisfaction of the need for autonomy, with each construct measured using three items. The item cited above represents an example of an autonomy satisfaction item. An example of a dissatisfaction item is: “When I used the study app, I felt a lot of pressure that I would have preferred to avoid” (German: “Wenn ich die Studienapp benutzte, habe ich viel Druck gespürt, auf den ich lieber verzichtet hätte”). Participants rated each statement on a 7-point Likert scale, indicating the extent to which it applied to them over the past month (ranging from “not at all” to “completely”). A composite score reflecting the overall degree of autonomy need fulfillment was calculated by subtracting the total dissatisfaction score from the total satisfaction score (Sheldon & Hilpert, 2012).

The possible range of scores thus spanned from -18 (e.g., ratings of 1 on all satisfaction items and 7 on all dissatisfaction items) to +18 (e.g., ratings of 7 on all satisfaction items and 1 on all dissatisfaction items). Scores from -18 to -6 were categorized as low perceived autonomy, scores from -5 to +5 as moderate perceived autonomy, and scores from +6 to +18 as high perceived autonomy.

Cronbach's alpha was calculated for the adapted autonomy scale to assess internal consistency. For this analysis, data from all groups were combined to provide an overall estimate of reliability. The resulting internal consistency (Cronbach's alpha) was $\alpha = .75$, which is considered acceptable based on conventional benchmarks ($\alpha > .70$ = acceptable; $\alpha > .80$ = good; $\alpha > .90$ = excellent) (Namdeo & Rout, 2016). The original version of the Balanced Measure of Psychological Needs Scale (Neubauer & Voss, 2016) demonstrated acceptable internal consistency, and the test-retest reliability was moderate. Specifically, the autonomy satisfaction subscale showed a test-retest correlation of $r_{tt} = .48$ and Cronbach's alpha values of .70 at t1 and .68 at t2. Similarly, the autonomy dissatisfaction subscale yielded a test-retest correlation of $r_{tt} = .47$, with internal consistency values of .61 at t1 and .73 at t2.

The construct validity of the autonomy subscales was supported by significant correlations with indicators of well-being and ill-being, with autonomy significantly predicting life satisfaction (Neubauer & Voss, 2016). These findings suggest that the adapted autonomy scale used in this study demonstrates a comparable level of reliability to the original version and can be considered a psychometrically sound instrument for assessing perceived autonomy in the given context.

Following the assessment of perceived autonomy, a combination of Likert scale items and open-ended questions was used to ask participants whether they were able to adhere to the tracking schedule.

Measurement of adherence. Two measures of adherence were used: absolute adherence and relative adherence between weekdays and weekends. Absolute adherence was required for testing Hypotheses 2a, 2b, 4, and 5a, while relative adherence was used to test Hypothesis 2c.

Absolute adherence was measured as the absolute frequency of tracking days on which participants followed their individual tracking plan as scheduled, without reference to the exact number of food items recorded on each allocated tracking day. The specific number of times participants logged their food intake each day was not considered relevant since a day was considered as successfully tracked if at least one entry was made. Relative adherence was calculated by dividing the number of days on which participants completed tracking as scheduled by the total number of days they were expected to track, based on their individual tracking plan.

To assess differences in adherence between weekdays and weekends, relative adherence was calculated separately for both time periods: weekdays (Monday to Friday) and weekends (Saturday and Sunday). The resulting difference score was obtained by subtracting weekend adherence from weekday adherence. Positive values indicate higher adherence during weekdays, whereas negative values suggest higher adherence on weekends.

Data Analysis

Data cleaning & screening

In this study, participants were included in the statistical analysis only if they tracked their fruit and vegetable portions on at least one scheduled day, thereby demonstrating a minimum level of adherence to the tracking schedule. If participants logged their tracking on days that were not part of their schedule, those days were excluded from the adherence analysis. Food tracking entries were also reviewed for plausibility. For instance, entries that incorrectly recorded items such as chocolate as a portion of fruit were excluded from further analysis. No participant was excluded due to insufficient reasonableness.

Out of the 168 participants who were randomized and signed up to m-path, only one participant from Group 3 was excluded from all analyses due to failing an instructional manipulation check. This person did not show one instance of adherence to the tracking protocol, indicating a lack of understanding of the task. The exclusion of this case resulted in a final sample of N = 167 for further analysis.

In three cases, participants did not specify the types of fruits and vegetables consumed but instead provided information about their preparation methods, such as "raw", “juice” or "steamed." These entries were retained, as the preparation terms imply the consumption of fruits and vegetables, suggesting that the participants likely understood the study's overall objective.

Of the 167 participants who were included for further analysis, nine did not complete the post-questionnaire. Consequently, only 158 participants were included in the data analyses for hypotheses H1, H3, H4, H5b and as these required post-questionnaire data. However, all 167 participants were included in the analyses of hypotheses H2a, H2b, H2c and H5a, which did not depend on post-questionnaire responses. Accordingly, pairwise deletion was applied.

Descriptive statistics

Descriptive statistics were used to characterize the sample both overall and separately for each of the three experimental groups (Groups 1 = self-selected schedule; Group 2 = weekdays-only; Group 3 = randomly assigned). Sociodemographic information included age (reported as mean, standard deviation, minimum, and maximum), gender (percentage of participants identifying as female, male, or diverse), highest level of education (percentages for the following categories: no school or training qualification, secondary school leaving certificate, intermediate school leaving certificate, Matura/high school diploma, apprenticeship/training, university degree, doctorate, and other), place of residence (percentage distribution across Austria, Germany, and other), and occupation (percentages for student, employed full-time, employed part-time, unemployed, retired, and other). Anthropometric characteristics were reported in terms of mean, standard deviation, minimum, and maximum for BMI (calculated as weight in kilograms divided by height in meters squared) and weight change from pre- to post-questionnaire, both overall and by group.

Randomization checks were conducted to ensure that the three groups did not differ significantly in demographic variables. As specified in the preregistration, *t*-tests were planned for metric variables (BMI and age) and chi-square tests were planned for categorical variables (gender, employment status, educational attainment, place of residence). However, during data analysis, it became evident that a one-way ANOVA would have been more appropriate for comparing the metric variables across three groups. Consequently, the assumptions of normality of residuals and homogeneity of variances were tested for these variables (See Table 2). Normality of residuals was assessed using Shapiro-Wilk tests, and Levene’s tests were used to assess homogeneity of variances. For age, Levene’s test indicated no significant differences in variances between groups ($p = .450$). However, residuals in all three groups deviated significantly from normality ($p < .001$). A similar pattern was observed for BMI. Levene’s test showed no significant variance differences between groups ($p = .976$), while residuals again showed significant deviations from normality in all groups ($p < .001$). For conscientiousness, Levene’s test indicated no significant differences in variances between groups ($p = .440$). However, residuals in all three groups deviated significantly from normality (Group 1: $p = .001$; Group 2: $p = .010$; Group 3: $p < .001$). Given the violations of the normality assumption, Kruskal-Wallis tests were used instead of ANOVAs to assess group differences in the metric variables. All assumptions for the Kruskal-Wallis test were met by the study design: independence of observations, an at least ordinally scaled dependent variable, independent groups, and similarly shaped distributions, which were confirmed through visual inspection of boxplots and histograms (see Figure 2).

Similarly, although chi-square tests were initially planned for the categorical variables, Fisher's exact tests were used instead due to the presence of categories with cell counts below five in each variable, which violates the assumptions of the chi-square test (see Table 3). Fisher's Exact tests do not require prior assumption testing since the assumptions of independence of observations and categorical variables are met by the study design.

With regard to primary outcome variables, absolute adherence was reported as mean and standard deviation for each group. This analysis related to hypotheses H2, H2a, H2b, H4, and H5a. In addition, relative adherence on weekdays vs. weekends was analyzed separately within Group 3 with the randomly assigned schedule to address Hypothesis 2c.

Study efficacy, defined as the change in average daily fruit and vegetable intake over the past week, was assessed by calculating the difference between pre- and post-intervention questionnaire responses. Means and standard deviations were reported both overall and separately for each group. Conscientiousness was described by its mean and standard deviation and was linked to Hypotheses H5a and H5b. Perceived autonomy was analyzed for each group, with mean and standard deviation reported in relation to Hypothesis 1.

Table 3

Sociodemographic Frequencies for Each Group (N = 167)

		Group 1 self-selected		Group 2 weekdays-only		Group 3 randomly assigned	
		n	%	*n*	%	*n*	%
Gender	male	15	25.9	10	18.5	16	29.0
	female	42	72.4	43	79.6	28	69.1
	non-binary	1	1.7	1	1.9	1	1.8
Recruitment	LABS	44	75.9	43	79.6	49	89.1
	other	14	24.1	11	20.4	6	10.9
Occupational status	student	49	84.5	44	81.5	49	89.1
	working full-time	4	6.9	4	7.4	2	3.6
	working part-time	0	0.0	3	5.6	2	3.6
	retired	2	3.5	2	3.7	0	0.0
	other	3	5.2	1	1.9	2	3.6
Level of education	highschool diploma	42	72.4	39	72.2	47	85.5
	university degree	14	24.1	14	25.9	7	12.7
	apprenticeship	2	3.5	1	1.9	0	0.0
	other	0	0.0	0	0.0	1	1.8

Inferential statistics

RQ1: How do the different self-monitoring schedules influence perceived autonomy over the course of the program?

To test Hypothesis 1, a between-subjects ANOVA was initially planned, with self-monitoring schedules (Groups 1, 2, and 3) as the independent variable and perceived autonomy as the dependent variable. Since the assessment of autonomy relied on data from the post-questionnaire, only participants who completed it were included in the analysis. Consequently, 9 out of the original 167 participants were excluded, resulting in a final sample of 158 (see Figure 1).

Prior to the analysis, statistical assumptions were examined (See Table 5). Independence of observations and continuous scaling of the dependent variable were ensured by the study design. Normality of residuals was assessed using a Shapiro-Wilk test, and homoscedasticity was evaluated with a Levene test. While Levene's test was not significant (p = .421), the Shapiro-Wilk test indicated a significant deviation from normality in all groups (Group 1: p = .032; Group 2: p = .006; Group 3: p = .012).

Due to the violation of the normality assumption, a nonparametric Kruskal-Wallis test was conducted instead of the planned ANOVA. All assumptions for the Kruskal-Wallis test were met by the study design: independence of observations, an at least ordinally scaled dependent variable, independent groups, and similarly shaped distributions, which were confirmed through visual inspection of boxplots and histograms (see Figure 3). Effect size was estimated using Eta² (η^2), with commonly used benchmarks for interpretation: values around .01 indicate a small effect, around .06 a medium effect, and .14 or above a large effect (Ellis, 2010).

RQ2: Do the different self-monitoring schedules differ in the adherence rate?

To test the hypothesis that self-monitoring schedules differ in terms of participant adherence (H2), a one-way ANOVA was planned. The independent variable was the type of self-monitoring schedule (three groups; nominal scale), and the dependent variable was adherence, measured as absolute frequency (interval scale). Since the analysis of Research Question 2 did not rely on data obtained through the post-questionnaire, the full sample of 167 participants was included in the analysis.

Prior to the main analysis, statistical assumptions were examined (see Table 6). Homogeneity of variances was assessed using a Levene test, which indicated no significant differences between groups (p = .390). Normality of the residuals was evaluated using the Shapiro-Wilk test, revealing significant deviations from normality in all groups (p < .001).

Due to violations of the normality assumption, a non-parametric Kruskal-Wallis test was conducted instead of a one-way ANOVA. All assumptions for the Kruskal-Wallis test were met by the study design: independence of observations, an at least ordinally scaled dependent variable, independent groups, and similarly shaped distributions, which were confirmed through visual inspection of boxplots and histograms (see Figure 5).

Effect size was estimated using Eta-Squared (η^2), with commonly used benchmarks for interpretation: values around .01 indicate a small effect, around .06 a medium effect, and .14 or above a large effect (Ellis, 2010).

Regardless of the overall test result, Dunn-Bonferroni post-hoc comparisons were performed to obtain specific contrast estimates for the directional hypotheses H2a and H2b. In addition, a one-sided Wilcoxon Signed-Rank Test was conducted within Group 3 to test Hypothesis 2c, examining whether relative adherence was higher on weekdays compared to weekends in the group with the randomly assigned schedule. Effect size was estimated using Pearson's correlation coefficient r, with commonly used benchmarks for interpretation: values around .10 indicate a small effect, around .30 a medium effect, and .50 or above a large effect (Ellis, 2010).

RQ3: Do different self-monitoring schedules lead to increased fruit / vegetable consumption?

To test Hypothesis 3, a mixed-design ANOVA was originally planned. The difference in fruit and vegetable consumption (study efficacy) was the dependent variable, measured on an interval scale at two time points: pre-intervention and post-intervention. Time (pre vs. post) served as the within-subject factor, while group, representing the assigned self-monitoring plans (nominally scaled), was included as the between-subject factor. The interaction term (Time × Group) was intended to assess whether changes in consumption over time differed across groups. Since the calculation of the difference in fruit and vegetable consumption (study efficacy) needed post-questionnaire data, only participants completing the post-questionnaire were included (N = 158).

Before conducting the main analysis, statistical assumptions were tested (see Table 7). Levene's test indicated that the assumption of homogeneity of variances was met (p = .786). However, the Shapiro-Wilk test revealed significant deviations from normality in Group 1 with the self-selected schedules and Group 2 with the weekday-only schedules, but not in Group

3 with the randomly assigned schedule (Group 1: p = .011; Group 2: p = .040; Group 3: p = .189). Because the normality assumption was violated, the Aligned Rank Transform (ART) ANOVA replaced the planned mixed-design ANOVA as a robust alternative, as its assumptions of independent observations, an at-least-ordinal measurement scale of the dependent variable, and homogeneity of variances were satisfied. Effect size was estimated using partial Eta² (partial η_p^2), with commonly used benchmarks for interpretation: values around .01 indicate a small effect, around .06 a medium effect, and .14 or above a large effect (Ortmann, 2023).

RQ4: Is there a correlation between study efficacy and adherence, and does this correlation vary in the different trial groups? Does conscientiousness moderate the impact of different food tracking schedules on both study efficacy and adherence?

To test Hypothesis 4, which assumes a positive relationship between adherence (IV) and study efficacy (DV) regardless of trial group (ModV), a hierarchical moderated regression analysis was planned. In practice, two separate moderated regressions were conducted: one comparing Group 1 (with the self-selected schedules) to Group 3 (with the randomly assigned schedules) and the other comparing Group 2 (with the weekdays-only schedule) to Group 3, with Group 3 consistently serving as the reference category (coded as 0). Since the calculation of study efficacy needed post-questionnaire data, only participants completing the post-questionnaire were included (N = 158; n = 106 for Group 1 vs 3; n = 105 for Group 2 vs 3).

Prior to conducting the moderated regression analysis, the statistical assumptions were examined (see Table 9 & 10). Linearity between the independent variable (adherence) and the dependent variable (study efficacy) was assessed using the Restricted Cubic Splines combined with a Likelihood Ratio Test (RCS + LRT). The comparison between the linear and the non-linear model resulted in a non-significant difference for the group comparison of Group 1 and Group 3 (p = .897), as well as for Group 2 and Group 3 (p = .608), suggesting that the assumption of linearity was met in both comparisons. The normality of residuals was evaluated using the Shapiro-Wilk test and revealed no significant deviations from normality for either group comparison (Group 1 vs 3: p = .200; Group 2 vs 3: p = .431). These results indicate that the assumption of normally distributed residuals was met. Homoscedasticity was tested using the Breusch-Pagan test. The results were not significant, indicating that the assumption of equal variance of residuals was met in both comparisons (Group 1 vs 3: p = .403; Group 2 vs 3: p = .063).

Multicollinearity between the centered independent variable (adherence) and the moderator variable (group) was assessed using the Variance Inflation Factor (VIF). VIF values ranged from 1.03 to 1.43 for the comparison between Group 1 and Group 3, and from 1.00 to 2.08 for the comparison between Group 2 and Group 3, suggesting that multicollinearity was not a concern in either case. The independence of residuals was examined using the Durbin-Watson test. The results were non-significant, indicating no violation of the independence assumption (Group 1 vs 3: p = .115; Group 2 vs 3: p = .387).

Additivity of the predictor variables was evaluated by examining the significance of the interaction term between the independent variable (adherence) and the moderator (trial groups) using a Wald test. The interaction was not statistically significant in either comparison, indicating that the assumption of additivity was met. Additivity was supported by p-values of p = .240 for the Group 1 and 3 comparison and p = .580 for the Group 2 and 3 comparison.

Since all assumptions for conducting a moderated regression are met, a hierarchical moderated regression analysis was deemed appropriate for testing Hypothesis 4 in the comparison between Group 1 (with the self-selected schedules) and Group 3 (with the randomly assigned schedules) as well as in the comparison between Group 2 (with the weekdays-only schedules) and Group 3. The hierarchical moderated regression model included the centered adherence variable, the group variable, and the interaction term as predictors. A statistically significant interaction would indicate that the relationship between adherence and study efficacy is moderated by the assigned trial group. Effect size was estimated using Cohen's f^2, with commonly used benchmarks for interpretation: values around .02 indicate a small effect, around .15 a medium effect, and .35 or above a large effect (Ellis, 2010).

RQ5: Does conscientiousness moderate the impact of different food tracking schedules on both study efficacy and adherence?

Hypothesis 5a. To test Hypothesis 5a, which assumes that conscientiousness moderates the influence of the different food tracking plans on adherence, with participants with higher conscientiousness showing higher adherence, a hierarchical moderated regression analysis was planned to be conducted.The independent variable (IV) were the trial groups (nominally scaled), and the dependent variable (DV) was adherence, measured as absolute frequency (interval-scaled). The moderator variable was conscientiousness (interval-scaled).

The procedure for the moderated regression analysis of Hypothesis 5a was identical to that for Hypothesis 4. Again two group-specific models were fitted. As before one comparison was between Group 1 (with the self-selected days) and Group 3 (with the randomly assigned days). While the other comparison was between Group 2 (with the weekdays-only schedules) and Group 3.

All 167 participants were included in the analysis. Based on the dummy coding, 113 cases were used in the comparison between Group 1 and Group 3, and 109 cases in the comparison between Group 2 and Group 3. Before calculating the interaction term, the moderator variable was mean-centered. The interaction term was then calculated by multiplying the centered conscientiousness and the group values.

Prior to conducting the moderated regression analysis, the statistical assumptions were examined (see Table 12 & 13). Linearity between the independent variable (trial groups) and the dependent variable (adherence) was assessed using the Restricted Cubic Splines combined with a Likelihood Ratio Test (RCS + LRT). The comparison between the linear and the non-linear model resulted in a significant difference for the group comparison of Group 1 and Group 3 (p = .018), indicating a deviation from linearity. Similarly, for the comparison between Group 2 and Group 3, the result was also significant (p = .017), again suggesting that the assumption of linearity was not met. The normality of residuals was evaluated using the Shapiro-Wilk test and revealed significant deviations from normality for either group comparison (p < .001). These results indicate that the assumption of normally distributed residuals was not met in both cases.

Homoscedasticity was tested using the Breusch-Pagan test. The results were not significant, indicating that the assumption of equal variance of residuals was met in both comparisons (Group 1 vs 3: p = .592; Group 2 vs 3: p = .087).

Multicollinearity between the centered independent variable (group) and the centered moderator variable (conscientiousness) was assessed using the Variance Inflation Factor (VIF). VIF values ranged from 1.00 to 1.75 for the comparison between Group 1 and Group 3, and from 1.00 to 1.92 for the comparison between Group 2 and Group 3, suggesting that multicollinearity was not a concern in either case. The independence of residuals was examined using the Durbin-Watson test. The results were non-significant, indicating no violation of the independence assumption (Group 1 vs 3: p = .565; Group 2 vs 3: p = .163).

Additivity of the predictor variables was evaluated by examining the significance of the interaction term between the independent variable (trial groups) and the moderator (conscientiousness) using a Wald test. The interaction was not statistically significant in either comparison, indicating that the assumption of additivity was met (Group 1 vs 3: p = .706; Group 2 vs 3: p = .052).

All assumptions held except for linearity and normally distributed residuals. To address linearity, scatterplots (see Figure 7) were inspected and showed an approximately straight-line relationship in both the comparison between Group 1 vs Group 3 and Group 2 vs Group 3 comparisons, thereby satisfying the linearity assumption. Because the residuals still deviated from normality, a bias-corrected bootstrapped regression with a resampling of 5,000 was used to obtain robust confidence intervals for the moderated regression, as this method does not rely on the assumption of normally distributed residuals. Consequently, a bootstrapped moderated regression was deemed appropriate for testing Hypothesis 5a in both group comparisons.

Hypothesis 5b. To test Hypothesis 5b, which assumes that conscientiousness moderates the influence of the different food tracking plans on adherence, with participants with higher conscientiousness showing higher adherence, a moderated regression analysis was planned to be conducted.The independent variable (IV) were the trial groups (nominally scaled), and the dependent variable (DV) was study efficacy (interval-scaled). The moderator variable was conscientiousness (interval-scaled).

As for Hypothesis 4 and Hypothesis 5a, two distinct moderated regression analyses were performed: one comparing Group 1 (with the self-selected schedules) to Group 3 (with teh randomly assigned schedules) and the other one comparing Group 2 (with the weekdays-only schedules) to Group 3. Group 3 was consistently used as the reference category (coded as 0). Because the assessment of study efficacy relied on data from the post-questionnaire, only participants who completed it were eligible for inclusion in the analysis. Consequently, 9 out of the original 167 participants were excluded, resulting in a final sample of 158 (see Figure 1). Due to dummy coding, 106 participants were included in the comparison between Group 1 and Group 3, and 105 participants in the comparison between Group 2 and Group 3. Before calculating the interaction term, the moderator variable was mean-centered. The interaction term was then calculated by multiplying the centered ModV (conscientiousness) with the IV (trial groups).

Prior to conducting the moderated regression analysis, the statistical assumptions were examined (See Table 15 & 16). Linearity between the independent variable (trial groups) and the dependent variable (study efficacy) was assessed using the Restricted Cubic Splines combined with a Likelihood Ratio Test (RCS + LRT). The comparison between the linear and the non-linear model resulted in a non-significant difference for the group comparison of Group 1 and Group 3 (p = .622), indicating no substantial deviation from linearity. Similarly, for the comparison between Group 2 and Group 3, the result was also non-significant (p = .276), again suggesting that the assumption of linearity was met.

The normality of residuals was evaluated using the Shapiro-Wilk test and revealed no significant deviations from normality for either group comparison (Group 1 vs 3: p = .058; Group 2 vs 3: p = .276). These results indicate that the assumption of normally distributed residuals was met in both cases. Homoscedasticity was tested using the Breusch-Pagan test. The results were not significant, indicating that the assumption of equal variance of residuals was met in both comparisons (Group 1 vs 3: p = .659; Group 2 vs 3: p = .669).

Multicollinearity between the centered independent variable (IV) and the centered moderator variable (ModV) was assessed using the Variance Inflation Factor (VIF). VIF values ranged from 1.00 to 1.71 for the comparison between Group 1 and Group 3, and from 1.01 to 1.94 for the comparison between Group 2 and Group 3, suggesting that multicollinearity was not a concern in either case. The independence of residuals was examined using the Durbin-Watson test. The results were non-significant, indicating no violation of the independence assumption for either comparison (Group 1 vs 3: p = .067; Group 2 vs 3: p = .076).

Additivity of the predictor variables was evaluated by examining the significance of the interaction term between the independent variable (trial groups) and the moderator (conscientiousness) using a Wald test. The interaction was not statistically significant in either comparison (Group 1 vs 3: p = .411; Group 2 vs 3: p = .056), indicating that the assumption of additivity was met.

Since all assumptions for conducting a moderated regression are met, a moderated regression analysis was deemed appropriate for testing Hypothesis 5b in the comparison between Group 1 and Group 3 as well as in the comparison between Group 2 and Group 3. Effect size was estimated using Cohen’s f^2, with commonly used benchmarks for interpretation: values around .02 indicate a small effect, around .15 a medium effect, and .35 or above a large effect (Ellis, 2010).

Results

Descriptive analysis

Descriptive analyses of the whole sample (*N* = 167) indicated that the sample was relatively young (*M* = 24.81, *SD* = 10.22), although the age range spanned broadly from 18 to 77 years. In terms of gender, 123 participants identified as female (73.7%), 41 as male (24.6%), and three as non-binary (1.8%). As expected, participation in the intervention did not lead to a meaningful reduction in body weight (*M* = 0.03, *SD* = 2.18), which aligned with the predefined inclusion criterion of a minimum body mass index (BMI) of 21. The final sample exhibited a mean BMI in the healthy range (*M* = 24.42, *SD* = 5.10). Regarding educational attainment, 128 participants (76.7%) held an Austrian or German high school diploma, 35 (21.0%) reported a university degree, three (1.8%) had completed an apprenticeship, and one participant (0.6%) selected “other.” The sample was largely homogeneous in terms of residence, with 152 participants (91.0%) living in Austria and 15 (9.0%) in Germany. Most participants were students (n = 142, 85.0%), with ten participants (6.0%) employed full-time, five (3.0%) part-time, four (2.4%) retired, and six (3.6%) not specifying their occupation. The majority were recruited via LABS (n = 136, 81.4%), while the remaining 31 participants (18.6%) were recruited through alternative channels. Overall adherence was high (*M* = 17.39, *SD* = 3.44). In addition, the difference in relative adherence between weekdays and weekends was examined in Group 3 (with the randomly assigned schedules). In line with Hypothesis 1c, there was a difference between weekdays and weekends in Group 3, however the difference was minimal, with a mean of 0.04 (*SD* = 0.17).

Since study efficacy and autonomy were assessed using post-questionnaire data, only participants who completed the post-questionnaire were included in the respective descriptive statistics. Across all groups, study efficacy, measured as the difference in fruit and vegetable consumption post versus pre intervention, was visible (*M* = 0.65, *SD* = 1.31). Conscientiousness scores fell within the medium range (*M* = 3.54, *SD* = 0.76), according to the categorization defined in the Methods section. Autonomy was high, with a mean of 8.19 (*SD* = 6.01).

The success of the randomization procedure was evaluated using a series of between-group comparisons. For the metric variables (BMI, age and conscientiousness), Kruskal-Wallis tests were conducted, while Fisher’s exact tests were used for all categorical variables. The analyses revealed no significant differences between the groups for any of the variables: BMI, $\chi^2(2) = 0.15$, $p = .927$; age, $\chi^2(2) = 2.52$, $p = .284$; conscientiousness, $\chi^2(2) = 0.53$, $p =$

.768; gender, *p* = .777; place of residence, *p* = .590; occupational status, *p* = .543; and highest level of education, *p* = .196. The descriptive statistics for the different groups concerning these variables highlight the success of the randomization procedure (see Table 3 and Table 4).

Table 4

Means and Standard Deviations for Each Group (N = 167)

	Group 1 self-selected		Group 2 weekdays-only		Group 3 randomly assigned	
	M	*SD*	*M*	*SD*	*M*	*SD*
age	25.34	11.05	25.78	11.93	23.29	6.98
BMI	24.35	4.74	24.49	5.71	24.41	4.93
weight difference (Post - Pre intervention)	-0.08	2.55	-0.11	2.48	0.28	1.30
study efficacy	0.79	1.41	0.73	1.16	0.43	1.34
absolute adherence	17.67	2.84	17.13	4.30	17.35	3.08
conscientiousness	3.52	0.69	3.60	0.79	3.51	0.81
autonomy	6.85	6.15	9.33	5.54	8.42	6.17

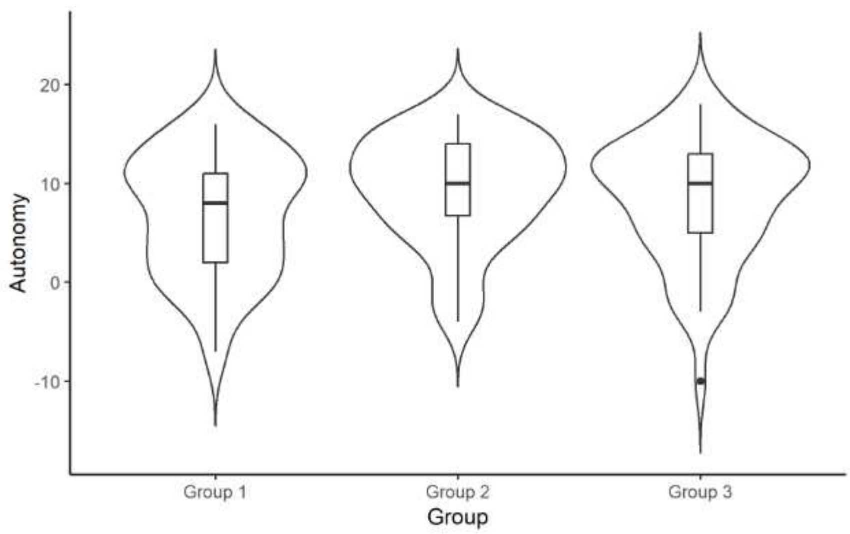
Primary outcomes

RQ1: How do the different self-monitoring schedules influence perceived autonomy over the course of the program?

As the assessment of autonomy relied on data obtained from the post-questionnaire, only participants who completed the post-questionnaire were included in the analysis. Therefore, the sample for the analysis of Research Question 1 consisted of 158 participants (see Figure 1). The Kruskal-Wallis test revealed no significant differences in perceived autonomy across the three groups, $H(2) = 4.31$, $p = .116$, which is also reflected in the small effect size of $\eta^2 = .03$ (See Figure 4). Consequently, Hypothesis 1, that Group 1 with the self-selected schedule would report higher autonomy than Groups 2 and 3, was not supported.

Figure 4

Violin Plot of Autonomy by Group



RQ2: Do the different self-monitoring schedules differ in the adherence rate?

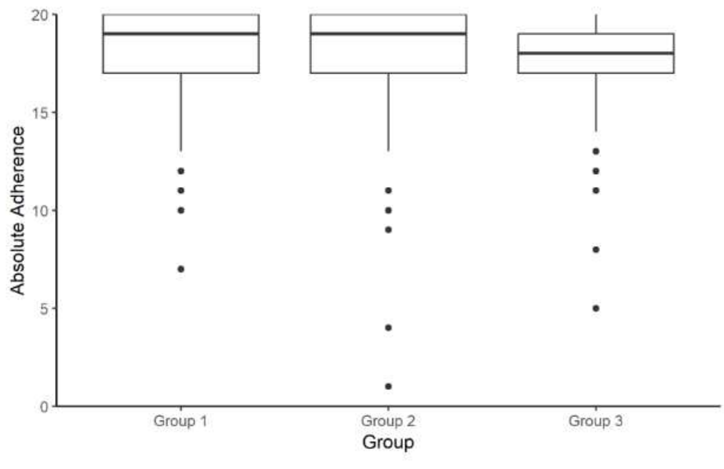
Concerning the hypothesis that self-monitoring schedules differ in terms of participant adherence (H2), a Kruskal-Wallis test was conducted to compare the effect of the different self-monitoring schedules (self-selected, weekdays-only, randomly assigned) on adherence. As Research Question 2 did not draw on data obtained from the post-questionnaire, all 167 participants were retained for this analysis.

There was no statistically significant main effect of tracking schedule on adherence ($H(2) = 1.42$, $p = .492$), which is also reflected in the small effect size of $\eta^2 = .01$, indicating that the groups did not differ in their adherence (see Figure 6). Despite the non-significant result of the overall Kruskal-Wallis test, Dunn-Bonferroni post-hoc comparisons were conducted to obtain specific estimates for the hypotheses 2a and 2b. Hypothesis 2a was tested by comparing Group 1 (with the self-selected schedule) and Group 3 (with the randomly assigned schedule), which yielded no significant difference, $Z = 0.803$, $p = .633$. Likewise Hypothesis 2b was tested by comparing Group 2 (with the weekdays-only schedules) to Group 3, which as well showed no significant difference, $Z = 1.164$, $p = .367$. Therefore also the specific group comparisons reflected the overall result that the groups did not differ in their adherence.

To test the directional Hypothesis 2c, examining whether relative adherence was higher on weekdays compared to weekends, a one-sided Wilcoxon Signed-Rank Test was conducted within Group 3. The result was not statistically significant ($V = 490.5$, $p = .080$), indicating that there was also no difference in relative adherence on weekdays compared to weekends in Group 3. This is also reflected in the effect size, which was estimated using Pearson's correlation coefficient, yielding $r = .225$, which corresponds to a small effect.

Figure 6

Boxplots of Absolute Adherence by Group



RQ3: Do different self-monitoring schedules lead to increased fruit / vegetable consumption?

Concerning Hypothesis 3, an Aligned Rank Transform (ART) ANOVA was conducted to examine the effects of time (pre vs. post) and group assignment on fruit and vegetable consumption (See Table 8). Because the change in fruit and vegetable consumption was based on both pre- and post-questionnaire data, only those participants who also completed the post-questionnaire were included (*N* = 158).

There was a statistically significant main effect of the within-subject factor Time *F*(1, 156) = 35.55, *p* < .001, with a large effect size (η□² = .103). This indicates that fruit and vegetable consumption significantly differed between the two measurement points, regardless of group membership. As visible in Table 4, fruit and vegetable consumption increased in all three trial groups with an overall mean change of *M* = 0.65 (*SD* = 1.31).

There was no statistically significant main effect of the between-subject factor Group, *F*(2, 156) = 0.97, *p* = .379, with a small effect size (η□² = .006), indicating that the three groups did not differ significantly in their overall consumption levels, averaged across time points. The interaction effect between group and time was not statistically significant, *F*(2, 310) = 1.08, *p* = .341, indicating that the pattern of change over time did not differ significantly between the groups. This suggests that all self-monitoring schedules were similarly effective in increasing the fruit and vegetable consumption of the participants, supporting the assumption underlying Hypothesis 3.

Table 8

Aligned Rank Transform ANOVA for Hypothesis 3 (N = 158)

Predictor	df	F	p	η^2□
Group	2, 310	0.97	.380	.006
Time	1, 310	35.55	< .001	.103
Group × Time	2, 310	1.08	.341	.007

RQ4: Is there a correlation between study efficacy and adherence, and does this correlation vary in the different trial groups?

To test Hypothesis 4, which stated that adherence would predict study efficacy independent of the trial group, two hierarchical moderated regression analyses were conducted, comparing Group 1 versus Group 3 and Group 2 versus Group 3 (see Table 11). In each analysis, adherence and the relevant group dummy were entered in step one, followed by their interaction term in step two. Of the original 167 participants, 158 provided complete data and were included in the analyses. Descriptive statistics for adherence and study efficacy by trial group are shown in Table 4.

In the comparison of Group 1 (with the self-selected schedules) and Group 3 (with the randomly assigned schedules), adherence and the group term accounted for 3.2 % of the variance in study efficacy, $R^2 = .032$, adjusted $R^2 = .013$. Introducing the adherence × Group interaction in Step 2 yielded a non-significant increase in explained variance ($\Delta R^2 = .013$, $p = .240$), and a small effect size, Cohen's $f^2 = .05$. In the full model, adherence was positively but not significantly related to study efficacy ($B = 0.10$, $p = .093$); the main effect of Group 1 was non-significant ($B = 0.34$, $p = .218$) and the interaction term also failed to reach significance ($p = .240$). In summary, adherence showed a positive but non-significant association with study efficacy, and neither the main effect of Group nor the adherence × Group interaction produced a further significant increase in explained variance.

An analogous analysis comparing Group 2 (with the weekdays-only schedules) and Group 3 produced similar results. In Step 1, centered adherence and the Group term explained 5.6 % of variance, $R^2 = .056$, adjusted $R^2 = .038$. Adding the adherence × Group interaction in Step 2 resulted in a negligible $\Delta R^2 = .003$ ($p = .580$) and a small effect size Cohen's $f^2 = .06$. In the final model adherence again showed a positive but non-significant effect ($B = 0.10$, $p = .063$), the Group main effect was non-significant ($p = .302$) as was the interaction term ($p = .580$). In summary, the Group 2 vs. Group 3 comparison likewise showed that adherence had a positive but non-significant relationship with study efficacy ($B = 0.10$, $p = .063$), and neither the main effect of Group 2 nor the adherence × Group interaction produced a meaningful or significant increase in explained variance ($\Delta R^2 = .003$, $p = .580$).

Although the regression coefficient B of adherence was positive in both cases, indicating a direct proportional relationship,across both analyses, the main effect of adherence was not significant, indicating that the relationship between adherence and study efficacy was not meaningful. Furthermore neither Group 1 nor Group 2 moderation effects were statistically significant, indicating that trial-group membership did not significantly alter the association between adherence and study efficacy.

Table 11

Hierarchical Moderated Regression Analyses for Hypothesis 4 (N = 158)

Panel A. Group 1 vs. Group 3 (N = 106)

Step	Predictor	*B*	*SE*	β	*t*	*p*	*F*Δ(1,102)
1	Adherence	0.10	0.06	.20	1.70	.093	
	Group	0.34	0.27	.12	1.24	.218	
2	Adherence × Group	-0.14	0.12	-.14	-1.18	.240	1.40

Panel B. Group 2 vs. Group 3 (N = 105)

Step	Predictor	*B*	*SE*	β	*t*	*p*	*F*Δ(1,102)
1	Adherence	0.10	0.06	.26	1.88	.063	
	Group	0.26	0.25	.11	1.04	.302	
2	Adherence × Group	-0.04	0.08	-.08	-0.55	.580	0.31

Note. In each panel, Step 1 enters the centered adherence and the dummy-coded group contrast; Step 2 adds the corresponding adherence × group interaction. *F*Δ refers to the change from the preceding step. Δ*p* is not reported separately, as it is identical to the p-value from step 2.

RQ5: Does conscientiousness moderate the impact of different food tracking schedules on both study efficacy and adherence?

Hypothesis 5a. To test Hypothesis 5a, which posited that conscientiousness moderates the relationship between trial group and adherence, a hierarchical moderated regression analysis was conducted using bias-corrected and accelerated bootstrapping with 5,000 resamples to estimate 95% confidence intervals (BCa CIs) (see Table 14). Two separate analyses using dummy-coded group comparisons (Group 1 vs. Group 3 and Group 2 vs. Group 3) were performed. In each analysis, conscientiousness (mean-centered) and the respective dummy-coded group variables were entered in Step 1, followed by the conscientiousness × group interaction terms in Step 2. All 167 participants were included in the overall model, while the specific group comparisons included 113 participants for Group 1 vs. Group 3, and 109 participants for Group 2 vs. Group 3, respectively. Descriptive statistics for conscientiousness and adherence across the experimental groups are provided in Table 4.

In Step 1 of the comparison between Group 1 (with the self-selected schedules) and Group 3 (with the randomly assigned schedules) the predictors were entered without the interaction term. Conscientiousness and the group term accounted for 0.4 % of the variance in adherence, $R^2 = .004$. Introducing the conscientiousness × Group interaction in Step 2 yielded a non-significant change in explained variance, ($\Delta R^2 = .001$, $p = .706$), and a negligible effect size of $f^2 = .001$. In the final model, neither the main effect of group ($p = .551$) nor the main effect of conscientiousness ($p = .608$) nor their interaction ($p = .706$), were significant predictors of adherence.

In the Group 2 (weekdays-only schedule) and Group 3 comparison, the predictors entered in Step 1 accounted for 1.6% of the variance ($R^2 = .016$). Introducing the conscientiousness × group interaction term in Step 2 slightly increased explained variance ($\Delta R^2 = .035$, $p = .052$), with a small effect size of $f^2 = .04$. In the final model, neither the main effect of group ($p = .675$) nor the main effect of conscientiousness ($p = .677$) were significant predictors of adherence. Only the interaction term approached significance ($p = .052$).

Overall, these findings provide only partial support for Hypothesis 5a. When rounded to two decimal figures the interaction term would count as significant. But since the interaction term of the Group 1 vs. Group 3 comparison was not even close to significance; it can be neglected. Therefore the final conclusion is that the relationship between trial group and adherence is not moderated by conscientiousness.

Table 14

Bootstrapped Hierarchical Moderated Regression Analyses for Hypothesis 5a (N = 167)

Panel A. Group 1 vs. Group 3 (N = 113)						
Step	Predictor	*B*	*SE*	95% CI [a]	*p*	*F*Δ(1,109)
1	Group	0.34	0.56	[-0.72, 1.52]	.551	
	Conscientiousness	-0.26	0.50	[-1.32, 0.78]	.608	
2	Group × Conscientiousness	0.29	0.76	[-1.43, 1.77]	.706	0.14
Panel B. Group 2 vs. Group 3 (N = 109)						
Step	Predictor	*B*	*SE*	95% CI [a]	*p*	*F*Δ(1,105)
1	Group	-0.30	0.71	[-1.77, 1.09]	.675	
	Conscientiousness	-0.26	0.62	[-1.34, 0.71]	677	
2	Group × Conscientiousness	1.75	0.89	[0.12, 3.71]	.052	3.87

Note. In each panel, Step 1 enters the centered adherence and the dummy-coded group contrast; Step 2 adds the corresponding adherence × group interaction. *F*Δ refers to the change from the preceding step. Δ*p* is not reported separately, as it is identical to the p-value from step 2.

[a] bias-corrected and accelerated bootstrapping with 5,000 resamples to estimate 95% confidence intervals (BCa CIs).

Hypothesis 5b. To test Hypothesis 5b, which proposed that conscientiousness predicts study efficacy independent of trial assignment, a hierarchical moderated regression analysis was performed (See Table 17). Specifically, two dummy-coded comparisons were conducted: one contrasting Group 1 with Group 3, and another contrasting Group 2 with Group 3. Additionally, an overall model incorporating both contrasts simultaneously was examined.In each analysis, centered conscientiousness and the relevant group dummy variable were entered in the first step, and the conscientiousness × group interaction term was entered in the second step. Of the 167 participants, 158 provided complete post-intervention data and were retained for analysis. The Group 1 (self-selected schedules) versus Group 3 (randomly assigned schedules) comparison included 106 participants and the Group 2 (weekdays-only schedule) versus Group 3 comparison included 105 participants. Descriptive statistics for conscientiousness and study efficacy by trial group are presented in Table 4.

In the comparison of Group 1 (self-selected schedules) and Group 3 (randomly assigned schedules), conscientiousness and the group term accounted for 17.3 % of the variance in study efficacy, $R^2 = .173$. Introducing the conscientiousness × Group interaction in Step 2 yielded a non-significant increase in explained variance ($\Delta R^2 = .006$, $p = .411$), with a small effect size, $f^2 = .04$. In the full model, conscientiousness was positively but not significantly related to study efficacy ($B = 0.33$, $p = .159$). The main effect Group was also non-significant ($p = .200$) as was the interaction term ($p = .411$). In summary, conscientiousness showed a positive but non-significant association with study efficacy ($B = 0.33$, $p = .159$), and the interaction produced no significant increase in explained variance.

An analogous analysis comparing Group 2 (weekdays-only schedules) and Group 3 (randomly assigned schedules) produced similar results. In Step 1, centered conscientiousness and the Group 2 dummy explained 1.5 % of variance, $R^2 = .015$. Adding the conscientiousness × Group 2 interaction in Step 2 resulted in $\Delta R^2 = .035$ ($p = .056$) with a small effect size of $f^2 = .05$. In the final model, conscientiousness again showed a positive but non-significant effect ($B = 0.33$, $p = .120$). The Group main effect was non-significant ($p = .211$) as was the interaction term ($p = .056$). In summary, the Group 2 versus Group 3 comparison likewise showed that conscientiousness had a positive but non-significant relationship with study efficacy and neither the main effect Group nor the interaction produced a meaningful or significant increase in explained variance. Across all analyses, neither the main effects of trial group nor the conscientiousness-by-group interactions were statistically significant, indicating that conscientiousness does not alter the association between trial group and study efficacy.

Table 17

Hierarchical Moderated Regression Analyses for Hypothesis 5b (N = 158)

Panel A. Group 1 vs. Group 3 (N = 106)

Step	Predictor	*B*	*SE*	β	*t*	*p*	*F*Δ(1,102)
1	Group	0.34	0.27	.13	1.29	.200	
	Conscientiousness	0.33	0.23	.18	1.42	.159	
2	Group × Conscientiousness	-0.30	0.36	-.11	-0.83	.411	0.68

Panel B. Group 2 vs. Group 3 (N = 105)

Step	Predictor	*B*	*SE*	β	*t*	*p*	*F*Δ(1,102)
1	Group	0.31	0.24	.12	1.26	.211	
	Conscientiousness	0.33	0.21	.21	1.57	.120	
2	Group × Conscientiousness	-0.59	0.31	-.26	-1.94	.056	1.70

Note. In each panel, Step 1 enters the centered conscientiousness and the dummy-coded group contrast; Step 2 adds the corresponding adherence × group interaction. *F*Δ refers to the change from the preceding step. Δ*p* is not reported separately, as it is identical to the p-value from step 2.

Discussion

The present study investigated how three self monitoring schedules (self-selected days, weekdays-only and randomly assigned tracking days), affect food tracking adherence, influence study efficacy (defined as increased fruit and vegetable consumption), moderate the impact of conscientiousness and influence perceived autonomy.

Summary

The analysis of Hypothesis 1 indicated that autonomy did not differ meaningfully between the trial groups. Similarly, contrary to the assumptions of Hypotheses 2, 2a, and 2b, no significant differences in adherence were observed across the groups. Even within the group with randomly assigned tracking days (Group 3), adherence did not differ substantially between weekdays and weekends, which contradicts the assumption underlying Hypothesis 2c. An exception was the analysis of Hypothesis 3 which confirmed that all trial groups showed an increase in fruit and vegetable consumption. Nevertheless, in contrast to Hypothesis 4, adherence was not meaningfully associated with study efficacy in any group. Conscientiousness showed a small moderating effect on the adherence difference between Group 2 and Group 3, but no such effect emerged when comparing Group 1 and Group 3 (H5a), nor did conscientiousness moderate the relationship between group assignment and study efficacy (H5b). Below, each research question and its underlying hypotheses are discussed under consideration of limitations, practical implications, and directions for future research.

Limitations, practical implications and directions for future research

RQ1: How do the different self-monitoring schedules influence perceived autonomy over the course of the program?

Contrary to Hypothesis 4, participants in the self-selected schedule group (Group 1) did not report higher perceived autonomy compared to the other groups. This finding suggests that the experimental manipulation was insufficient to effectively foster a sense of autonomy in Group 1. In other words, simply allowing participants to choose their tracking schedule at the beginning of the study may not have been experienced as a meaningful expression of self-determination.

To enhance the feeling of autonomy in future interventions, more dynamic designs should be considered. For instance, allowing participants to modify their tracking schedules throughout the study period could promote a sustained sense of autonomy. Additionally, offering deeper customization options, beyond scheduling alone could further induce the feeling of autonomy.

An additional point that warrants consideration is that all three groups reported relatively high levels of autonomy. Rather than focusing solely on ways to further increase autonomy scores through additional experimental manipulations, future research should also address the question of why participants in the fixed schedule conditions reported such high levels of perceived autonomy. Particularly in the control group where all tracking days were randomly assigned. This unexpected finding raises the question of whether other factors such as general study design, perceived voluntariness, or individual differences in the need for autonomy may have contributed to the elevated autonomy ratings across all groups.

RQ2: Do the different self-monitoring schedules differ in the adherence rate?

Contrary to Hypotheses 2a to 2c, neither allowing participants to choose their own tracking days (Group 1) nor restricting tracking to weekdays (Group 2) led to higher adherence compared to randomly assigned tracking days (Group 3). At first glance, these findings suggest that neither fostering autonomy through self-selection nor avoiding weekend monitoring meaningfully affected adherence. However, there are alternative explanations worth considering.

With respect to Hypothesis 2a, it is important to note that Hypothesis 1 (predicting higher perceived autonomy in Group 1) was not supported. As such, the underlying theoretical mechanism linking increased autonomy to greater adherence may not have been successfully activated. All groups reported similarly high autonomy scores, not just the self-selected schedule group as initially expected. Given that the autonomy manipulation appears to have been ineffective, it is plausible that adherence in Group 1 did not differ significantly from that in the control group.

Similarly, with respect to H2b and H2c, it is important to consider the uniformly high adherence observed across all three groups. Whereas studies by Pellegrini et al. (2018) and Krukowski et al. (2013) tracked participants over six months and Krukowski et al. (2013) reported average adherence rates of roughly 70-75 percent in the first month, present data yielded adherence rates of approximately 83-87 percent (weekend vs. weekday) across conditions. Several factors could account for this discrepancy. The shorter duration of the present study may have mitigated the fatigue typically seen over longer trials.

Moreover, the composition of the present sample likely drove adherence even higher. While the sample of Krukowski et al. (2013) consisted of a voluntary overweight cohort, this sample consists predominantly of psychology students earning credits for their participation. These students may have felt compelled to comply throughout the 30 days through fear that they may not receive full credits otherwise. In light of this, the lack of difference between the tracking schedules in general (H2) may also simply reflect a generally high level of commitment rather than true equivalence of tracking behavior across trial groups (H2a & H2b) or specific tracking days (H2c).

Another factor worth considering is the overall study design, which was intentionally structured to make the tracking experience as pleasant as possible. Findings from the systematic review by König et al. (2021) indicate that participants are often dissatisfied with food tracking apps, particularly those requiring extensive manual input and those focused on calorie counting. Users also reported frustration with overly frequent notifications (Lieffers et al., 2017). König et al. (2021) concluded that simple, low-effort tracking processes and high usability increase the likelihood of user adoption and continued engagement. In line with these insights, the present study deliberately employed a simple and user-friendly tracking process and refrained from sending notifications throughout the tracking period, except for a single reminder at the end to prompt completion of the post-questionnaire. Consequently, not only the characteristics of the sample but also the study design itself may have contributed to the generally high adherence levels observed across all groups.

RQ3: Do different self-monitoring schedules lead to increased fruit / vegetable consumption?

The results of the study confirmed Hypothesis 3 that all trial groups would experience an increase in fruit and vegetable consumption. The underlying theory that self-monitoring is a highly effective behavior change strategy was confirmed many times before (Berry et al., 2021; Villinger et al., 2019). Likewise, the results concerning Hypothesis 3 are in line with the results of Gardiner and Bryan (2021) who found that self-monitoring facilitates behavior change in eating habits, including increased fruit and vegetable consumption. In line with these findings, the present results also suggest that simplified food tracking approaches, as proposed by König et al. (2021), hold considerable potential. Prior research by Mandracchia et al. (2019) and Raber et al. (2021) similarly highlights the value of minimalistic tracking methods, showing that tracking fruit and vegetable intake alone can yield positive effects on health and weight management, increase dietary awareness, and reduce the burden of self-monitoring for participants. Therefore, maybe it is time to reinvent the structure of popular food tracking applications.

RQ4: Is there a correlation between study efficacy and adherence, and does this correlation vary in the different trial groups?

The anticipated positive link between adherence and study efficacy (i.e., increased fruit and vegetable intake) across all groups (H4) was not supported by the data. Given that the systematic review by Patel et al. (2021) included over fifty studies, 39 of which were randomized controlled trials, it is highly unlikely that the present study meaningfully contradicts their findings.

A more plausible explanation is that certain factors in the current study design prevented the expected relationship from emerging. One such factor could be the relatively short duration of the intervention. While the studies reviewed by Patel et al. (2021) spanned at least 12 weeks, the present study lasted only four weeks. Moreover, since both adherence and study efficacy were generally high across all participants, statistical variability was limited, making it more difficult to detect associations. This phenomenon, known as restriction of range, can attenuate correlation coefficients and reduce statistical power, thereby obscuring true relationships between variables (Sackett et al., 2007). In this case, the uniformly high adherence and efficacy scores may have resulted in insufficient variance to reveal a significant link between the two constructs, even if such a relationship exists under more diverse conditions. Future studies should therefore consider broader sampling and a longer tracking period that allow for greater variability in participant behavior to better assess such associations.

RQ5: Does conscientiousness moderate the impact of different food tracking schedules on both study efficacy and adherence?

In regards to H5a, which stated that, conscientiousness moderates the influence of the different food tracking plans on adherence, only a marginal effect emerged when comparing Group 2 to the random schedule group (Group 3). More conscientious participants tended to adhere slightly better when tracking was restricted to weekdays. However, this moderation effect did not reach full conventional significance in the separate contrast ($p = .052$), and no moderation emerged for Group 1 versus Group 3. Likewise the analysis of H5b did not support the underlying theory.

As for H4, the absence of a robust moderation effect by conscientiousness is likely attributable to the combination of consistently high adherence (H5a), high study efficacy (H5b) and high conscientiousness scores within the sample. This homogeneity restricted variability in the key variables reduces the sensitivity of the analysis to detect meaningful interactions. In statistical terms, a limited range in both the predictor and outcome variables may reduce observable effects, even if true associations exist (Sackett et al., 2007). In this context, the lack of variation may have concealed potential moderation effects that could emerge under more diverse conditions.

Therefore, future research should consider recruiting a more heterogeneous sample and extending the tracking period to enhance behavioral variability. Such design adjustments could provide a clearer picture of how personality traits like conscientiousness interact with structural features of self-monitoring interventions.

An alternative explanation, however, is that there may genuinely be no meaningful relationship between conscientiousness and adherence to food tracking interventions or an increased fruit and vegetable consumption (study efficacy). While previous research has linked conscientiousness to better adherence and study efficacy in other domains, the evidence is not without limitations. For instance, the study by Lakshmanan et al. (2024) found that more conscientious individuals adhered more consistently to exercise schedules. However, this study had a small sample size (N = 44) and exclusively included older adults (aged 57 to 95), which limits the generalizability of the findings to broader and younger populations. Furthermore, food tracking may rely on different motivational and cognitive processes than physical activity, making it inappropriate to generalize effects across domains without direct evidence.

Future studies should therefore not only refine methodological design but also distinguish between behavioral contexts to better understand how conscientiousness interacts with specific types of self-regulatory demands. Only then can we determine whether conscientiousness is a consistent predictor of adherence and study efficacy across various health-related behaviors, or whether its influence is domain-specific.

Conclusion

Overall, these null findings underscore the need for a more heterogeneous sample. Without sufficient variability among participants, even well-established effects from the literature can go undetected. Moreover do the results highlight how extrinsic incentives such as the possibility of earning credits through study participation can confound intervention effects and obscure the true relationship between variables. Furthermore the failure of the autonomy manipulation shows the importance of further research on which factors precisely lead to the general high levels of autonomy across all trial groups. And how alternative ways to induce the feeling of autonomy would affect the autonomy inducement. Future interventions should therefore eliminate extrinsic rewards, extend monitoring periods, and recruit more diverse samples to enhance generalizability and variance. By addressing these design considerations, future research can meaningfully shape the design of tracking apps and thus help promote healthier eating behaviors.

References

Bardus, M., Van Beurden, S. B., Smith, J. R. & Abraham, C. (2016). A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. *International Journal Of Behavioral Nutrition And Physical Activity*, *13*(1). https://doi.org/10.1186/s12966-016-0359-9

Berry, R., Kassavou, A. & Sutton, S. (2021). Does self‐monitoring diet and physical activity behaviors using digital technology support adults with obesity or overweight to lose weight? A systematic literature review with meta‐analysis. *Obesity Reviews*, *22*(10). https://doi.org/10.1111/obr.13306

Bogg, T. & Roberts, B. W. (2004). Conscientiousness and Health-Related Behaviors: A Meta-Analysis of the Leading Behavioral Contributors to Mortality. *Psychological Bulletin*, *130*(6), 887–919. https://doi.org/10.1037/0033-2909.130.6.887

Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz. (2024). *Österreichische Ernährungsempfehlungen NEU*. https://www.sozialministerium.at/Themen/Gesundheit/Ern%C3%A4hrung/%C3%96sterreichische-Ern%C3%A4hrungsempfehlungen-NEU.html (Abgerufen am 04.02.2025)

Burke, L., Acharya, N., Elci, N., Sereika, N., Music, N., Styn, N. & Turk, N. (2009). Adherence to a behavioral weight loss treatment program enhances weight loss and improvements in biomarkers. *Patient Preference And Adherence*, 151. https://doi.org/10.2147/ppa.s5802

Carver, C. S. & Scheier, M. F. (1998). *On the Self-Regulation of Behavior*. https://doi.org/10.1017/cbo9781139174794

Chopik, W. & Lee, J. H. (2022). Dyadic Associations Between Conscientiousness Facets, Health, and Health Behavior Over Time. *Collabra Psychology*, *8*(1). https://doi.org/10.1525/collabra.37611

Clark-Carter, D. (2007). Effect Size and Statistical Power in Psychological Research. *The Irish Journal Of Psychology*, *28*(1–2), 3–12. https://doi.org/10.1080/03033910.2007.10446244

Ellis, P. D. (2010). *The Essential Guide to Effect Sizes*. https://doi.org/10.1017/cbo9780511761676

Flaherty, S. J., McCarthy, M. B., Collins, A. M., & McAuliffe, F. M. (2019). A different perspective on consumer engagement: exploring the experience of using health apps to support healthier food purchasing. *Journal of Marketing Management*, *35*(3-4), 310-337. https://doi.org/10.1080/0267257X.2019.1576756

G*Power. (2020, 17. March). *G*Power.* https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower (Abgerufen am 04.12.2024)

Galindo-Perdomo, F., Peiró-Velert, C. & Valencia-Peris, A. (2021). Do Adolescents Who Meet Physical Activity Recommendations on Weekdays Also Meet Them on Weekends? A Cross-Sectional Study in Colombia. *International Journal Of Environmental Research And Public Health*, *18*(3), 897. https://doi.org/10.3390/ijerph18030897

Gardiner, C. K. & Bryan, A. D. (2021). Randomized controlled trials of self‐monitoring interventions with or without incentives for diet and exercise among individuals with overweight or obesity: Psychological and behavioural effects. *British Journal Of Health Psychology*, *26*(4), 1114–1134. https://doi.org/10.1111/bjhp.12523

Gartland, N., Wilson, A., Lawton, R. & O'Connor, D. B. (2020). Conscientiousness and engagement with national health behaviour guidelines. *Psychology Health & Medicine*, *26*(4), 421–432. https://doi.org/10.1080/13548506.2020.1814961

Glusman, M., Bruce, A., Thelen, J., Smith, J., Lynch, S., Catley, D., Bennett, K. K., & Bruce, J. (2019). Discontinuing Treatment Against Medical Advice: The role of perceived autonomy support from providers in relapsing-remitting multiple sclerosis. *Annals of Behavioral Medicine, 53*(3), 283–289. https://doi.org/10.1093/abm/kay035

Gold, J. M., Carr, L. J., Thomas, J. G., Burrus, J., O'Leary, K. C., Wing, R. & Bond, D. S. (2020). Conscientiousness in weight loss maintainers and regainers. *Health Psychology*, *39*(5), 421–429. https://doi.org/10.1037/hea0000846

Gubelmann, C., Marques-Vidal, P., Bringolf-Isler, B., Suggs, L. S., Vollenweider, P. & Kayser, B. (2017). Correlates of weekday compliance to physical activity recommendations in Swiss youth non-compliant in weekend days. *Preventive Medicine Reports*, *9*, 86–91. https://doi.org/10.1016/j.pmedr.2017.12.004

Haahr, M. (o. D.). *RANDOM.ORG - List Randomizer*. RANDOM.ORG. https://www.random.org/lists/ (Abgerufen am 04.02.2025)

Harvey, J., Krukowski, R., Priest, J., & West, D. (2019). Log often, lose more: Electronic dietary self-monitoring for weight loss. *Obesity*, *27*(3), 380–384. https://doi.org/10.1002/oby.22382

Hongu, N., Kataura, M. P. & Block, L. M. (2011). *Behavior Change Strategies for Successful Long-Term Weight Loss: Focusing on Dietary and Physical Activity Adherence, Not Weight Loss.* https://eric.ed.gov/?id=EJ917172

Inzlicht, M., Werner, K. M., Briskin, J. L. & Roberts, B. W. (2021). Integrating Models of Self-Regulation. *Annual Review Of Psychology*, *72*(1), 319–345. https://doi.org/10.1146/annurev-psych-061020-105721

Kaiser, K. A., Brown, A. W., Brown, M. M. B., Shikany, J. M., Mattes, R. D. & Allison, D. B. (2014). Increased fruit and vegetable intake has no discernible effect on weight loss: A systematic review and meta-analysis. *The American Journal of Clinical Nutrition, 100*(2), 567–576. https://doi.org/10.3945/ajcn.114.090548

Kennedy, S., Goggin, K., & Nollen, N. (2004). Adherence to HIV Medications: Utility of the Theory of Self-Determination. *Cognitive Therapy and Research, 28*(5), 611–628. https://doi.org/10.1023/B:COTR.0000045568.95219.e2

Kern, M. L. & Friedman, H. S. (2008). Do conscientious individuals live longer? A quantitative review. *Health Psychology*, *27*(5), 505–512. https://doi.org/10.1037/0278-6133.27.5.505

Kgasi, M. R., Chimbo, B. & Motsi, L. (2024). A persuasive technology mhealth self-monitoring system for intervention in diabetic patients medical adherence. *International Journal Of Science Annals*, *7*(2). https://doi.org/10.26697/ijsa.2024.2.2

König, L. M., Attig, C., Franke, T. & Renner, B. (2021). Barriers to and Facilitators for Using Nutrition Apps: Systematic Review and Conceptual Framework. *JMIR Mhealth And Uhealth*, *9*(6), e20037. https://doi.org/10.2196/20037

Krukowski, R. A., Harvey-Berino, J., Bursac, Z., Ashikaga, T. & West, D. S. (2013). Patterns of success: Online self-monitoring in a web-based behavioral weight control program. *Health Psychology*, *32*(2), 164–170. https://doi.org/10.1037/a0028135

Lakshmanan, S., Flynn, S., Reyes, J., Rubinstein, S. & Zelinski, E. (2024). I FOLLOW THROUGH: CONSCIENTIOUSNESS IS ASSOCIATED WITH ADHERENCE. *Innovation in Aging*, *8*(Supplement_1), 1318–1319. https://doi.org/10.1093/geroni/igae098.4211

Lee, L., Hall, R., Stanley, J. & Krebs, J. (2023). Tailored Prompting to Improve Adherence to Image-Based Dietary Assessment: Mixed Methods Study. *JMIR Mhealth And Uhealth*, *12*, e52074. https://doi.org/10.2196/52074

Lieffers, J. R., Arocha, J. F., Grindrod, K. & Hanning, R. M. (2017). Experiences and Perceptions of Adults Accessing Publicly Available Nutrition Behavior-Change Mobile Apps for Weight Management. *Journal Of The Academy Of Nutrition And Dietetics*, *118*(2), 229-239.e3. https://doi.org/10.1016/j.jand.2017.04.015

Michie, S., Abraham, C., Whittington, C., McAteer, J. & Gupta, S. (2009). Effective techniques in healthy eating and physical activity interventions: A meta-regression. *Health Psychology*, *28*(6), 690–701. https://doi.org/10.1037/a0016136

Mossavar-Rahmani, Y., Henry, H., Rodabough, R., Bragg, C., Brewer, A., Freed, T., Kinzel, L., Pedersen, M., Soule, C. & Vosburg, S. (2004). Additional self-monitoring tools in the dietary modification component of the women's health initiative. *Journal Of The American Dietetic Association*, *104*(1), 76–85. https://doi.org/10.1016/j.jada.2003.10.017

m-path. (n.d.). *m-path*. https://m-path.io/ (Abgerufen am 04.02.2025)

Namdeo, S. & Rout, S. (2016). Calculating and interpreting Cronbach's alpha using Rosenberg assessment scale on paediatrician's attitude and perception on self esteem. *International Journal Of Community Medicine And Public Health*, 1371–1374. https://doi.org/10.18203/2394-6040.ijcmph20161448

Neubauer, A. B., & Voss, A. (2016). Validation and revision of a German version of the Balanced Measure of Psychological Needs Scale. *Journal of Individual Differences, 37*(1), 56–72. https://doi.org/10.1027/1614-0001/a000188

Ng, J. Y. Y., Ntoumanis, N., Thøgersen-Ntoumani, C., Deci, E. L., Ryan, R. M., Duda, J. L. & Williams, G. C. (2012). Self-Determination Theory Applied to Health Contexts: A Meta-Analysis. *Perspectives On Psychological Science*, *7*(4), 325–340. https://doi.org/10.1177/1745691612447309

Ortmann, M. (2023). *Partielles EtA Quadrat | Dr. Magdalene Ortmann | Statistik für Mediziner*. Dr. Magdalene Ortmann. https://ortmann-statistik.de/glossar/partielles-eta-quadrat/ (Abgerufen am 13.06.2025)

Patel, M.L., Wakayama, L.N. and Bennett, G.G. (2021), Self-monitoring via digital health in weight loss interventions: A systematic review among adults with overweight or obesity. *Obesity*, 29: 478-499. https://doi.org/10.1002/oby.23088

Pellegrini, C. A., Conroy, D. E., Phillips, S. M., Pfammatter, A. F., McFadden, H. G. & Spring, B. (2018). Daily and seasonal influences on dietary self-monitoring using a smartphone application. *Journal of Nutrition Education and Behavior*, *50*(1), 56-61.e1. https://doi.org/10.1016/j.jneb.2016.12.004

Python Software Foundation. (n.d.). *Python*. https://www.python.org/ (Abgerufen am 04.02.2025)

Raber, M., Liao, Y., Rara, A., Schembre, S. M., Krause, K. J., Strong, L., Daniel-MacDougall, C. & Basen-Engquist, K. (2021). A systematic review of the use of dietary self-monitoring in behavioural weight loss interventions: delivery, intensity and effectiveness. Public Health Nutrition, 24(17), 5885–5913. https://doi.org/10.1017/s136898002100358x

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2012). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit : Big-Five-Inventory-10 (BFI-10). *Methoden, Daten, Analysen (Mda)*, *7*(2), 32. https://madoc.bib.uni-mannheim.de/54728/ (Abgerufen am 04.02.2025)

Rivera, J., McPherson, A., Hamilton, J., Birken, C., Coons, M., Iyer, S., Agarwal, A., Lalloo, C. & Stinson, J. (2016). Mobile Apps for Weight Management: A Scoping Review. *JMIR Mhealth And Uhealth*, *4*(3), e87. https://doi.org/10.2196/mhealth.5115

Rush, E. C., & Yan, M. R. (2017). Evolution not revolution: Nutrition and obesity. *Nutrients*, *9*(5), 519. https://doi.org/10.3390/nu9050519

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being*. American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sackett, P. R., Lievens, F., Berry, C. M. & Landers, R. N. (2007). A cautionary note on the effects of range restriction on predictor intercorrelations. *Journal Of Applied Psychology*, *92*(2), 538–544. https://doi.org/10.1037/0021-9010.92.2.538

Sheldon, K. M., & Hilpert, J. C. (2012). The balanced measure of psychological needs (BMPN) scale: An alternative domain general measure of need satisfaction. *Motivation and Emotion, 36*(4), 439–451. https://doi.org/10.1007/s11031-012-9279-4

Soto, C. J., Kronauer, A. & Liang, J. K. (2015). Five‐Factor Model of Personality. *The Encyclopedia Of Adulthood And Aging*, 1–5. https://doi.org/10.1002/9781118521373.wbeaa014

Statistik Austria. (2025). *Ernährung*. STATISTIK AUSTRIA. https://www.statistik.at/statistiken/bevoelkerung-und-soziales/gesundheit/gesundheitsverhalten/ernaehrung (Abgerufen am 04.02.2025)

Tivian XI GmbH. (2021). Unipark [Software]. https://www.unipark.com/umfragesoftware/ (Abgerufen am 04.02.2025)

Turner-McGrievy, G. M., Dunn, C. G., Wilcox, S., Boutté, A. K., Hutto, B., Hoover, A. & Muth, E. (2019). Defining Adherence to Mobile Dietary Self-Monitoring and Assessing Tracking Over Time: Tracking at Least Two Eating Occasions per Day Is Best Marker of Adherence within Two Different Mobile Health Randomized Weight Loss Interventions. *Journal Of The Academy Of Nutrition And Dietetics*, *119*(9), 1516–1524. https://doi.org/10.1016/j.jand.2019.03.012

Ushioda, E. (2011). Why autonomy? Insights from motivation theory and research. *Innovation in Language Learning and Teaching, 5*(2), 221–232. https://doi.org/10.1080/17501229.2011.577536

Vervloet, M., Spreeuwenberg, P., Bouvy, M. L., Heerdink, E. R., De Bakker, D. H. & Van Dijk, L. (2013). Lazy Sunday afternoons: the negative impact of interruptions in patients' daily routine on adherence to oral antidiabetic medication. A multilevel analysis of electronic monitoring data. *European Journal Of Clinical Pharmacology*, *69*(8), 1599–1606. https://doi.org/10.1007/s00228-013-1511-y

Villinger, K., Wahl, D. R., Boeing, H., Schupp, H. T. & Renner, B. (2019). The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: A systematic review and meta-analysis. *Obesity Reviews*, *20*(10), 1465–1484. https://doi.org/10.1111/obr.12903

Wilkinson, R. (2017). Rising spirals and virtuous circles: The interrelationship between motivation and learner autonomy. In R. Breeze & C. Sancho Guinda (Hrsg.), *Essential Competencies for English-medium University Teaching* (S. 219–232). Springer International Publishing. https://doi.org/10.1007/978-3-319-40956-6_15

World Health Organization. (2003). *Adherence to long-term therapies : evidence for action*. https://iris.who.int/handle/10665/42682 (Abgerufen am 04.02.2025)

List of Tables

Table 1

Deviations from the Preregistration

Topic of Deviation	Preregistration	Deviation	Explanation
Research Questions	RQ3: Is there a correlation between study efficacy and adherence, and does this correlation vary in the different trial groups? Does conscientiousness moderate the impact of different food tracking schedules on both study efficacy and adherence?	RQ3: Is there a correlation between study efficacy and adherence, and does this correlation vary in the different trial groups? RQ4: Does conscientiousness moderate the impact of different food tracking schedules on both study efficacy and adherence?	Since RQ3 actually contains two research questions, it was split into RQ3 and RQ4.
	RQ4: How do the different self-monitoring schedules influence perceived autonomy over the course of the program?	RQ1: How do the different self-monitoring schedules influence perceived autonomy over the course of the program?	This was previously RQ4 (the last research question). Since this RQ functions as a manipulation check for the autonomy inducement, it was moved to the beginning (now RQ1).

Randomization Check for continuous variables	covered only age and BMI	conscientiousness was included as well	The randomization check was extended to include conscientiousness to ensure comparability between the groups regarding this personality trait, which plays a central role in the research design and interpretation of results.
	t-tests for the analysis	Kruskal-Wallis tests were used instead	ANOVAs would have been more suitable for comparisons across three groups. After ANOVA assumption checks revealed violations of residual normality, Kruskal-Wallis tests were used instead.
Randomization Check for categorical variables	chi-square tests	Fisher's exact test	Fisher's exact test replaced chi-square tests due to violated normality and low cell counts.
Analysis of Hypothesis 1	Breusch-Pagan test was planned to test the assumption of homoscedasticity.	Levene test was used instead.	Levene test was conducted instead since it is the standard and more commonly used method for assessing equality of group variances when conducting an ANOVA.

Analysis of Hypothesis 3	post-hoc tests	no post-hoc tests	No post-hoc tests were conducted for the within-subject factor *Time*, as it included only two levels (pre and post), rendering pairwise comparisons unnecessary.
Analysis of Hypothesis 4, 5a and 5b	Moderated regressions were planned for the analysis of H4, H5a and H5b. The preregistration stated that all IV and ModV would be mean-centered. In Hypothesis 4, the moderator variable (ModV) was the trial group. In Hypotheses 5a and 5b, the independent variable (IV) was the trial group.	only metric variables were mean-centered (dummy-coded group variables were not)	According to standard procedure in moderation analysis, only continuous predictors were mean-centered. Dummy-coded variables representing group assignments were excluded from centering, as they are categorical by definition.
	The assumption of Linearity was intended to be analyzed with the Box-Tidwell test.	Linearity was assessed with restricted cubic splines plus likelihood ratio tests (RCS+LRT).	The Box-Tidwell test could not be used, given that predictors had values of zero and below.

Table 2

Assumption Testing of Homoscedasticity and Normality of Residuals for Randomization Checks

Variable	Test	Group	Statistic		df	*p*
			F	*W*		
Age	Levene's Test		0.80		2, 165	.450
	Shapiro-Wilk Test	Group 1		0.55		< .001
		Group 2		0.55		< .001
		Group 3		0.57		< .001
BMI	Levene's Test		0.03		2, 165	.976
	Shapiro-Wilk Test	Group 1		0.61		< .001
		Group 2		0.52		< .001
		Group 3		0.59		< .001
conscientiousness	Levene's Test		0.82		2, 165	.440
	Shapiro-Wilk Test	Group 1		0.93		.001
		Group 2		0.94		.010
		Group 3		0.90		< .001

Note. Due to significant deviations from normality in all groups, Kruskal-Wallis tests were used to assess group differences in the metric variables instead of one-way ANOVAs.The assumption of similarly shaped distributions for the Kruskal-Wallis test were confirmed through visual inspection of boxplots and histograms (see Figure 2).

Table 3

Sociodemographic Frequencies for Each Group (N = 167)

		Group 1 self-selected		Group 2 weekdays-only		Group 3 randomly assigned	
		n	%	*n*	%	*n*	%
Gender	male	15	25.9	10	18.5	16	29.0
	female	42	72.4	43	79.6	28	69.1
	non-binary	1	1.7	1	1.9	1	1.8
Recruitment	LABS	44	75.9	43	79.6	49	89.1
	other	14	24.1	11	20.4	6	10.9
Occupational status	student	49	84.5	44	81.5	49	89.1
	working full-time	4	6.9	4	7.4	2	3.6
	working part-time	0	0.0	3	5.6	2	3.6
	retired	2	3.5	2	3.7	0	0.0
	other	3	5.2	1	1.9	2	3.6
Level of education	highschool diploma	42	72.4	39	72.2	47	85.5
	university degree	14	24.1	14	25.9	7	12.7
	apprenticeship	2	3.5	1	1.9	0	0.0
	other	0	0.0	0	0.0	1	1.8

Table 4

Means and Standard Deviations for Each Group (N = 167)

	Group 1 self-selected		Group 2 weekdays-only		Group 3 randomly assigned	
	M	*SD*	*M*	*SD*	*M*	*SD*
age	25.34	11.05	25.78	11.93	23.29	6.98
BMI	24.35	4.74	24.49	5.71	24.41	4.93
weight difference (post - pre intervention)	-0.08	2.55	-0.11	2.48	0.28	1.30
study efficacy	0.79	1.41	0.73	1.16	0.43	1.34
absolute adherence	17.67	2.84	17.13	4.30	17.35	3.08
conscientiousness	3.52	0.69	3.60	0.79	3.51	0.81
autonomy	6.85	6.15	9.33	5.54	8.42	6.17

Table 5

Assumption Testing of Homoscedasticity and Normality of Residuals for the between-subjects ANOVA of Hypothesis 1

Test	Group	Statistic		df	*p*
		F	*W*		
Levene's Test		0.87		2, 156	.421
Shapiro-Wilk Test	Group 1		0.95		.032
	Group 2		0.93		.006
	Group 3		0.94		.012

Note. Due to significant deviations from normality in all groups, a Kruskal-Wallis test was used instead.The assumption of similarly shaped distributions for the Kruskal-Wallis test were confirmed through visual inspection of boxplots and histograms (see Figure 3).

Table 6

Assumption Testing of Homoscedasticity and Normality of Residuals for the one-way ANOVA Analysis of Hypothesis 2

Variable	Test	Group	Statistic		df	*p*
			F	*W*		
adherence	Levene's Test		0.95		2, 165	.390
	Shapiro-Wilk Test	Group 1		0.79		< .001
		Group 2		0.71		< .001
		Group 3		0.78		< .001

Note. Due to significant deviations from normality in all groups, a Kruskal-Wallis test was used instead of an one-way ANOVA.The assumption of similarly shaped distributions for the Kruskal-Wallis test were confirmed through visual inspection of boxplots and histograms (see Figure 5).

Table 7

Assumption Testing of Homoscedasticity and Normality of Residuals for the mixed-design ANOVA Analysis of Hypothesis 3

Test	Group	Statistic		df	*p*
		F	*W*		
Levene's Test		0.24		2, 156	.786
Shapiro-Wilk Test	Group 1		0.94		.011
	Group 2		0.95		.040
	Group 3		0.97		.189

Note. Due to significant deviations from normality in Group 1 and Group 2, an Aligned Rank Transform (ART) ANOVA was used instead of a mixed-design ANOVA.

Table 8

Aligned Rank Transform ANOVA for Hypothesis 3 (N = 158)

Predictor	df	F	p	η^2
Group	2, 310	0.97	.380	.006
Time	1, 310	35.55	< .001	.103
Group × Time	2, 310	1.08	.341	.007

Table 9

Assumption Tests for the moderated regressions of Hypothesis 4 (Group 1 vs Group 3)

Assumption	Test	Statistic	df	p
Linearity	RCS + LRT, χ^2	3.53	8	.897
Normality of residuals	Shapiro-Wilk, *W*	.98		.200
Homoscedasticity	Breusch-Pagan, χ^2	2.93	3	.403
Independence of errors	Durbin-Watson, *DW*	1.76		.115
Additivity	Wald Test, *t*	-1.18	102	.240

Note. RCS = Restricted Cubic Splines (Regression); LRT = likelihood-ratio test. Multicollinearity was evaluated through the Variance Inflation Factor (VIF) with values of 1.43 for the centered IV (absolute adherence), 1.03 for ModV (dummy coded group variable) as well as 1.42 for the interaction term.

Table 10

Assumption Tests for the moderated regressions of Hypothesis 4 (Group 2 vs Group 3)

Assumption	Test	Statistic	df	p
Linearity	RCS + LRT, χ^2	6.35	8	.608
Normality of residuals	Shapiro-Wilk, *W*	.99		.431
Homoscedasticity	Breusch-Pagan, χ^2	7.31	3	.063
Independence of errors	Durbin-Watson, *DW*	1.94		.387
Additivity	Wald Test, *t*	-.55	101	.580

Note. RCS = Restricted Cubic Splines (Regression); LRT = likelihood-ratio test. Multicollinearity was evaluated through the Variance Inflation Factor (VIF) with values of 2.08 for the centered IV (absolute adherence), 1.00 for ModV (dummy coded group variable) as well as 2.08 for the interaction term.

Table 11

Hierarchical Moderated Regression Analyses for Hypothesis 4 (N = 158)

Panel A. Group 1 vs. Group 3 (N = 106)

Step	Predictor	*B*	*SE*	β	*t*	*p*	*F*Δ(1,102)
1	Adherence	0.10	0.06	.20	1.70	.093	
	Group	0.34	0.27	.12	1.24	.218	
2	Adherence × Group	-0.14	0.12	-.14	-1.18	.240	1.40

Panel B. Group 2 vs. Group 3 (N = 105)

Step	Predictor	*B*	*SE*	β	*t*	*p*	*F*Δ(1,102)
1	Adherence	0.10	0.06	.26	1.88	.063	
	Group	0.26	0.25	.11	1.04	.302	
2	Adherence × Group	-0.04	0.08	-.08	-0.55	.580	0.31

Note. In each panel, Step 1 enters the centered adherence and the dummy-coded group contrast; Step 2 adds the corresponding adherence × group interaction. *F*Δ refers to the change from the preceding step. Δ*p* is not reported separately, as it is identical to the p-value from step 2.

Table 12

Assumption Tests for the moderated regressions of Hypothesis 5a (Group 1 vs Group 3)

Assumption	Test	Statistic	df	p
Linearity	RCS + LRT, χ^2	11.97	4	.018
Normality of residuals	Shapiro-Wilk, *W*	.81		< .001
Homoscedasticity	Breusch-Pagan, χ^2	1.91	3	.592
Independence of errors	Durbin-Watson, *DW*	2.02		.565
Additivity	Wald Test, *t*	.38	109	.706

Note. RCS = Restricted Cubic Splines (Regression); LRT = likelihood-ratio test. Multicollinearity was evaluated through the Variance Inflation Factor (VIF) with values of 1.00 for the group dummy variable, 1.75 for the centered ModV (conscientiousness) as well as 1.75 for the interaction term.

Table 13

Assumption Tests for the moderated regressions of Hypothesis 5a (Group 2 vs Group 3)

Assumption	Test	Statistic	df	p
Linearity	RCS + LRT, χ^2	10.21	3	.017
Normality of residuals	Shapiro-Wilk, *W*	.82		< .001
Homoscedasticity	Breusch-Pagan, χ^2	6.57	3	.087
Independence of errors	Durbin-Watson, *DW*	1.81		.163
Additivity	Wald Test, *t*	1.97	105	.052

Note. RCS = Restricted Cubic Splines (Regression); LRT = likelihood-ratio test. Multicollinearity was evaluated through the Variance Inflation Factor (VIF) with values of 1.00 for the group dummy variable, 1.92 for the centered ModV (conscientiousness) as well as 1.92 for the interaction term.

Table 14

Bootstrapped Hierarchical Moderated Regression Analyses for Hypothesis 5a (N = 167)

Panel A. Group 1 vs. Group 3 (N = 113)

Step	Predictor	*B*	*SE*	95% CI [a]	*p*	*F*Δ(1,109)
1	Group	0.34	0.56	[-0.72, 1.52]	.551	
	Conscientiousness	-0.26	0.50	[-1.32, 0.78]	.608	
2	Group × Conscientiousness	0.29	0.76	[-1.43, 1.77]	.706	0.14

Panel B. Group 2 vs. Group 3 (N = 109)

Step	Predictor	*B*	*SE*	95% CI [a]	*p*	*F*Δ(1,105)
1	Group	-0.30	0.71	[-1.77, 1.09]	.675	
	Conscientiousness	-0.26	0.62	[-1.34, 0.71]	677	
2	Group × Conscientiousness	1.75	0.89	[0.12, 3.71]	.052	3.87

Note. In each panel, Step 1 enters the centered adherence and the dummy-coded group contrast; Step 2 adds the corresponding adherence × group interaction. *F*Δ refers to the change from the preceding step. Δ*p* is not reported separately, as it is identical to the p-value from step 2.

[a] bias-corrected and accelerated bootstrapping with 5,000 resamples to estimate 95% confidence intervals (BCa CIs).

Table 15

Assumption Tests for the moderated regressions of Hypothesis 5b (Group 1 vs Group 3)

Assumption	Test	Statistic	df	p
Linearity	RCS + LRT, χ^2	2.63	4	.622
Normality of residuals	Shapiro-Wilk, *W*	.98		.058
Homoscedasticity	Breusch-Pagan, χ^2	1.60	3	.659
Independence of errors	Durbin-Watson, *DW*	1.71		.067
Additivity	Wald Test, *t*	-.83	102	.411

Note. RCS = Restricted Cubic Splines (Regression); LRT = likelihood-ratio test. Multicollinearity was evaluated through the Variance Inflation Factor (VIF) with values of 1.00 for the group dummy variable, 1.71 for the centered ModV (conscientiousness) as well as 1.71 for the interaction term.

Table 16

Assumption Tests for the moderated regressions of Hypothesis 5b (Group 2 vs Group 3)

Assumption	Test	Statistic	df	p
Linearity	RCS + LRT, χ^2	3.87	3	.276
Normality of residuals	Shapiro-Wilk, *W*	.99		.431
Homoscedasticity	Breusch-Pagan, χ^2	1.56	3	.669
Independence of errors	Durbin-Watson, *DW*	1.72		.076
Additivity	Wald Test, *t*	-1.94	101	.056

Note. RCS = Restricted Cubic Splines (Regression); LRT = likelihood-ratio test. Multicollinearity was evaluated through the Variance Inflation Factor (VIF) with values of 1.01 for the group dummy variable, 1.94 for the centered ModV (conscientiousness) as well as 1.93 for the interaction term.

Table 17

Hierarchical Moderated Regression Analyses for Hypothesis 5b (N = 158)

Panel A. Group 1 vs. Group 3 (N = 106)

Step	Predictor	*B*	*SE*	β	*t*	*p*	*F*Δ(1,102)
1	Group	0.34	0.27	.13	1.29	.200	
	Conscientiousness	0.33	0.23	.18	1.42	.159	
2	Group × Conscientiousness	-0.30	0.36	-.11	-0.83	.411	0.68

Panel B. Group 2 vs. Group 3 (N = 105)

Step	Predictor	*B*	*SE*	β	*t*	*p*	*F*Δ(1,102)
1	Group	0.31	0.24	.12	1.26	.211	
	Conscientiousness	0.33	0.21	.21	1.57	.120	
2	Group × Conscientiousness	-0.59	0.31	-.26	-1.94	.056	1.70

Note. In each panel, Step 1 enters the centered conscientiousness and the dummy-coded group contrast; Step 2 adds the corresponding adherence × group interaction. *F*Δ refers to the change from the preceding step. Δ*p* is not reported separately, as it is identical to the p-value from step 2.

List of Figures

Figure 1

Flow diagram of participant selection and drop outs

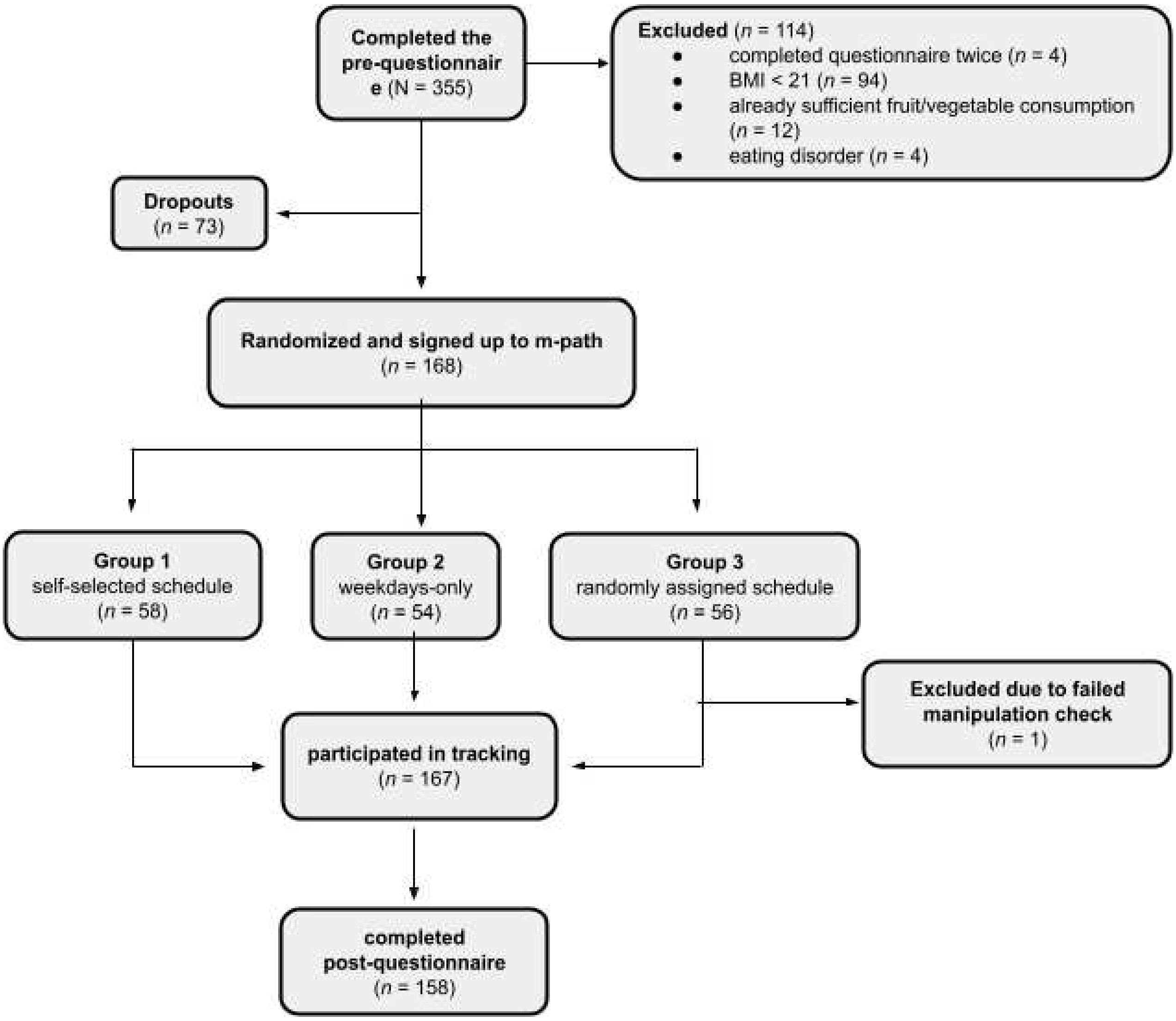
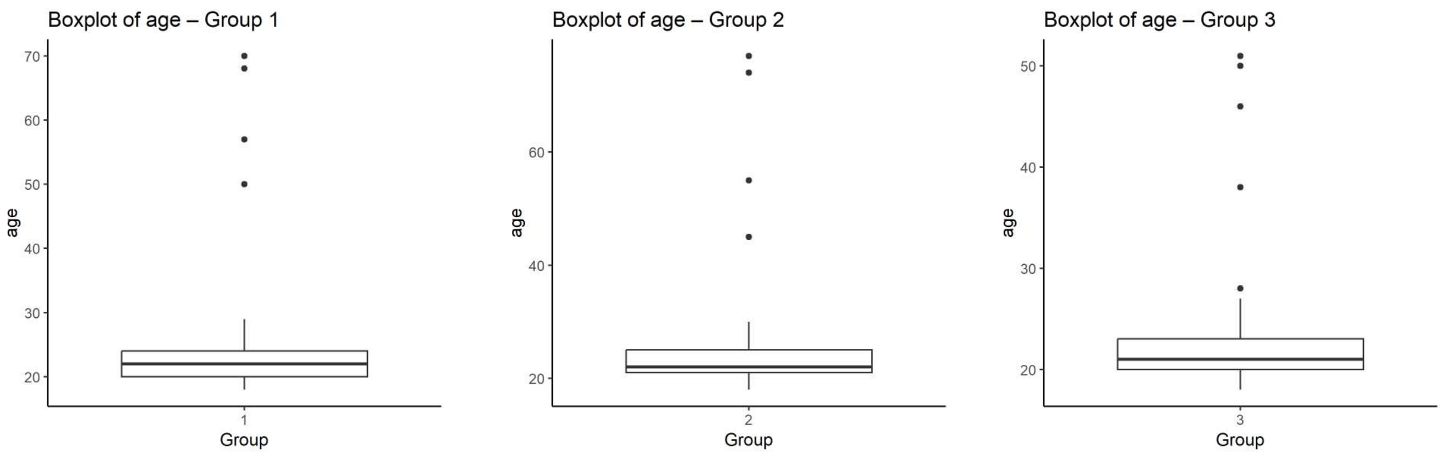
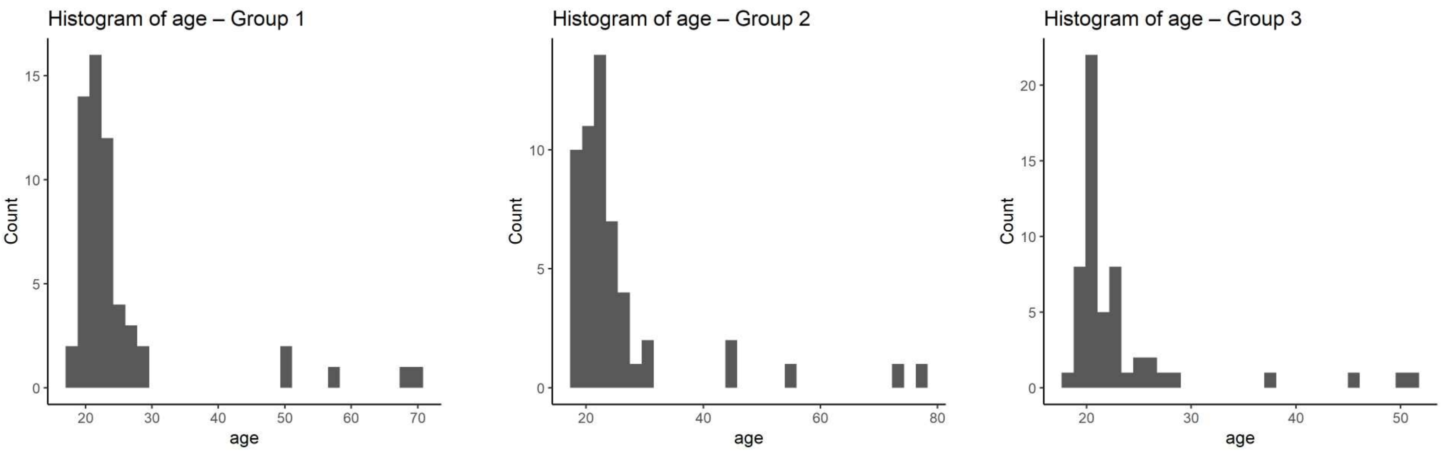


Figure 2

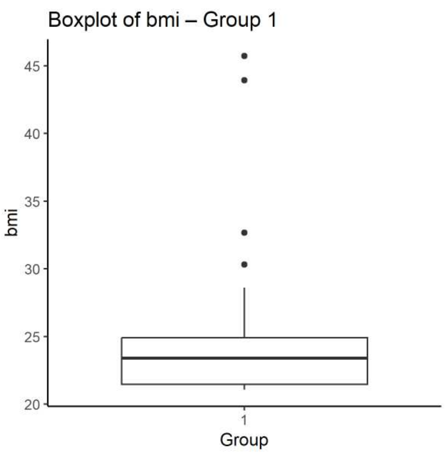
Visual assumption testing for the Kruskal-Wallis test of variables age, BMI and conscientiousness (Randomization Check)

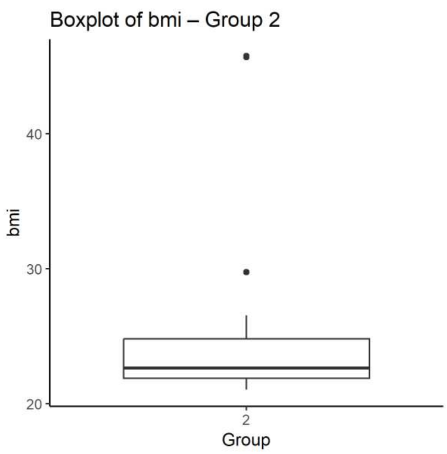


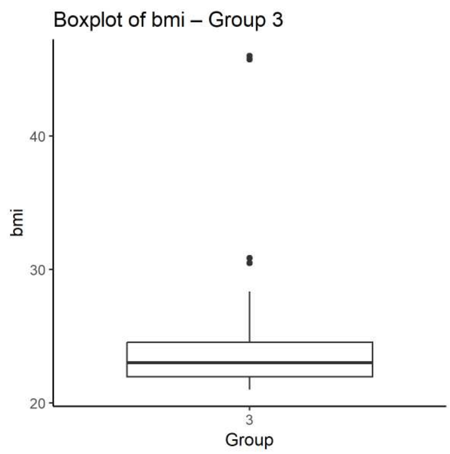
Variable: age

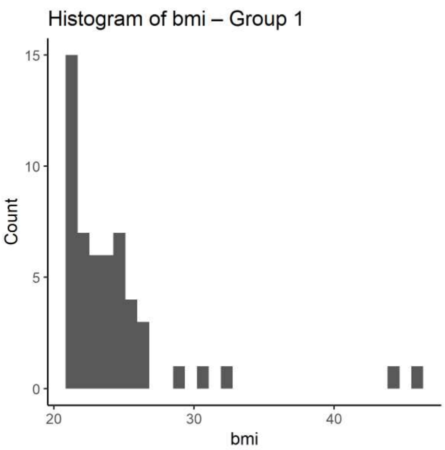


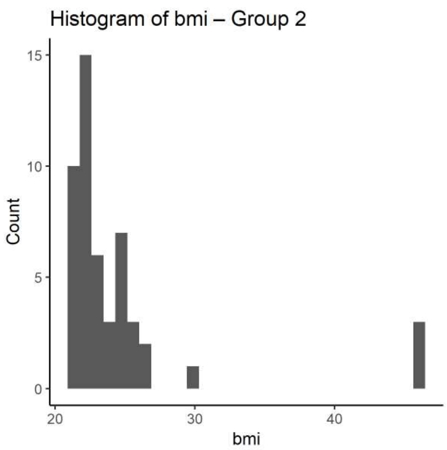
Variable: BMI

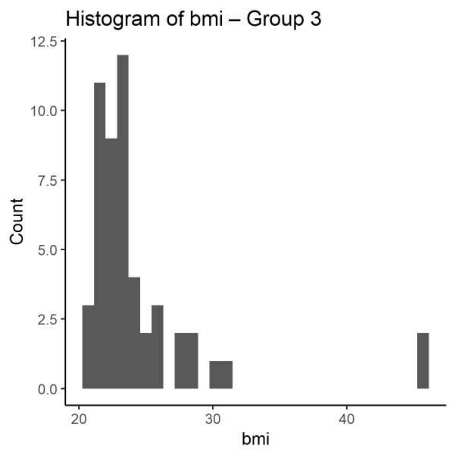




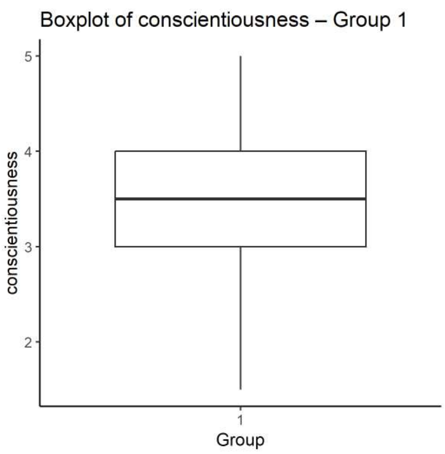


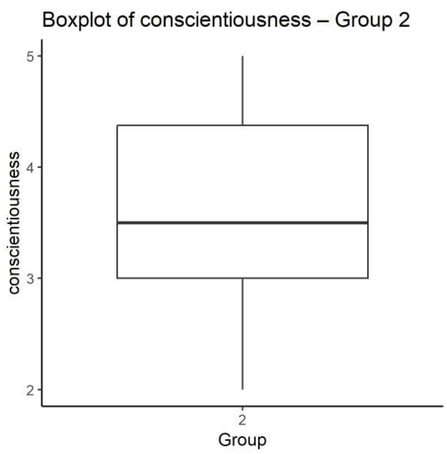


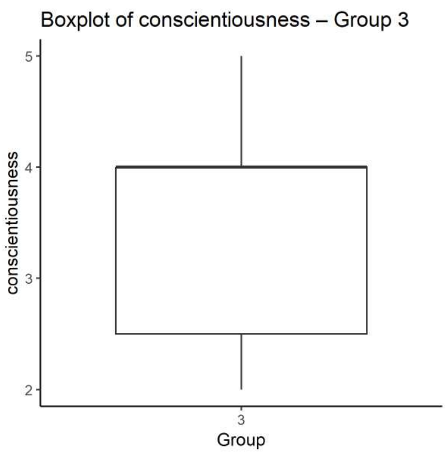


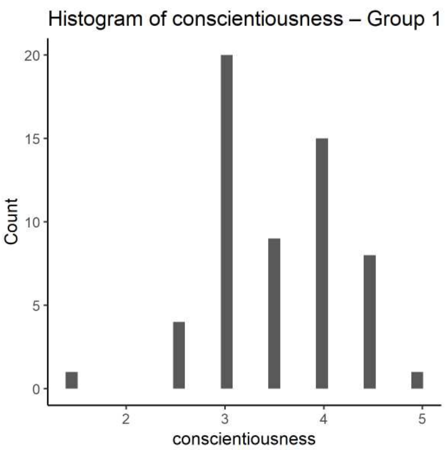


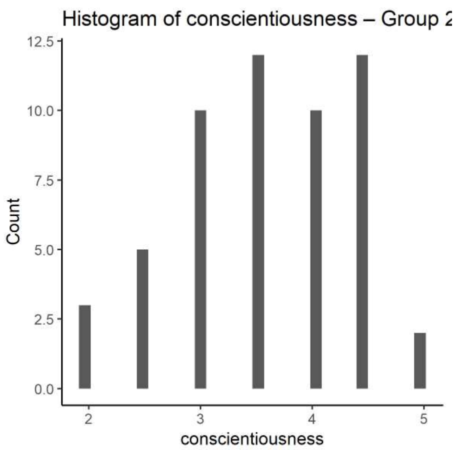
Variable: conscientiousness











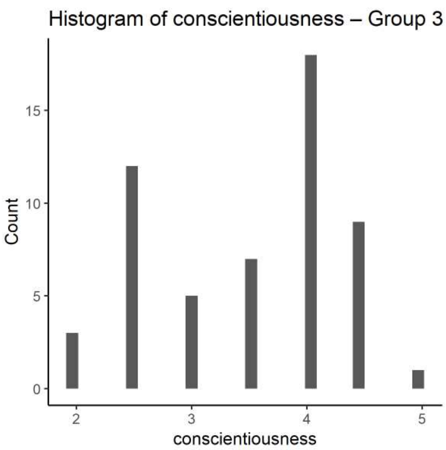
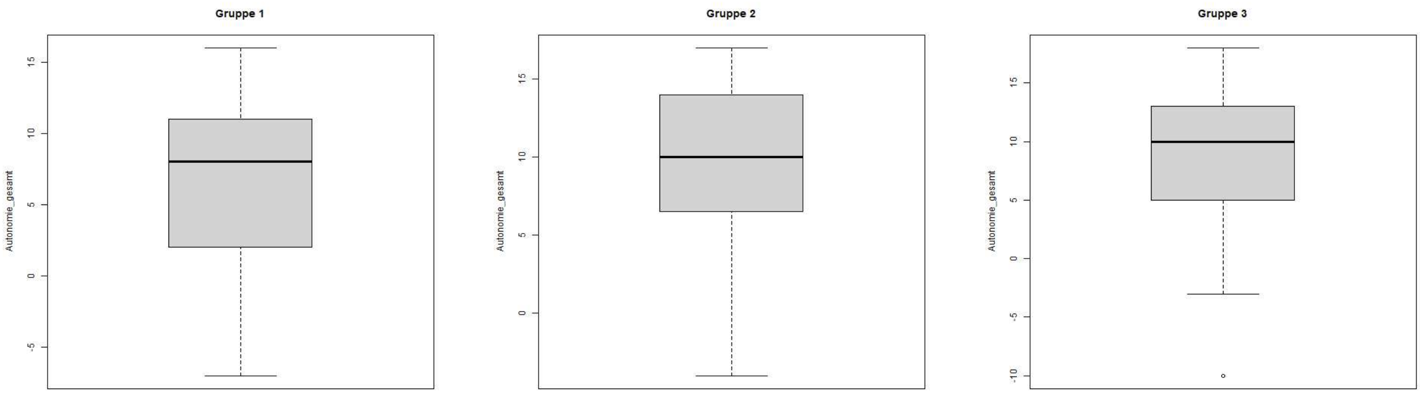


Figure 3

Visual assumption testing for the Kruskal-Wallis test - Boxplots and Histograms (Hypothesis 1)

Boxplots



Histograms

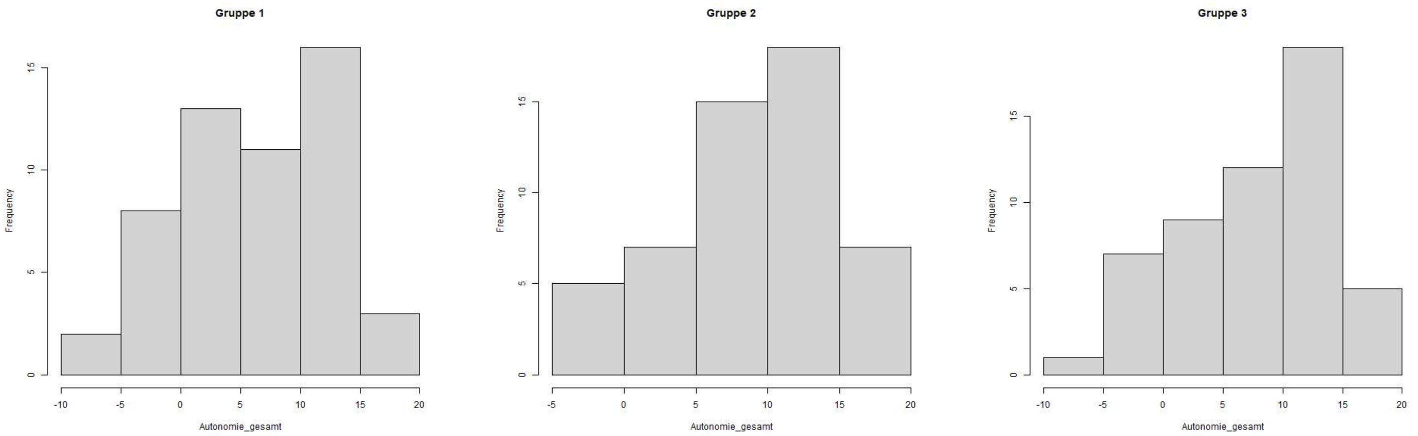


Figure 4

Violin Plot of Autonomy by Group (Hypothesis 1)

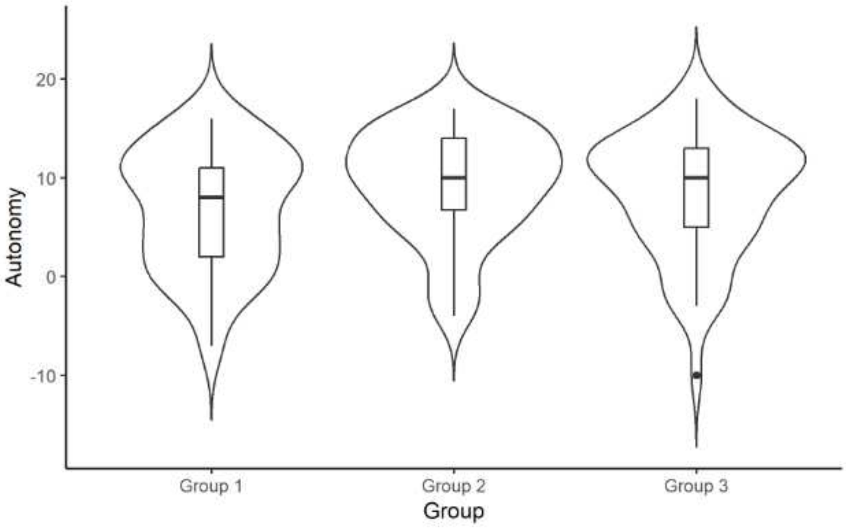
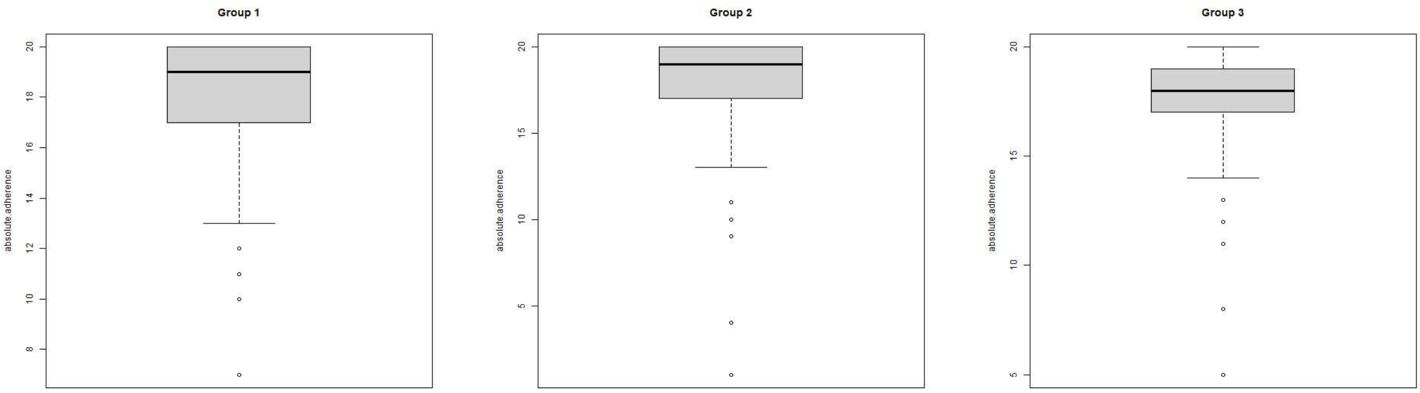


Figure 5

Visual assumption testing for the Kruskal-Wallis test - Boxplots and Histograms (Hypothesis 2)

Boxplots



Histograms

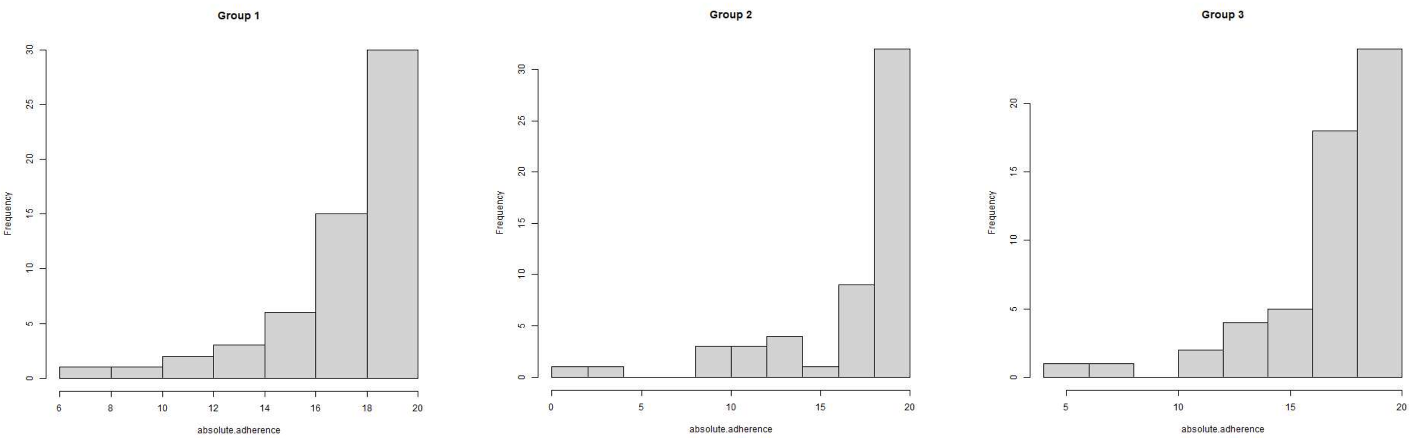


Figure 6

Boxplots of Absolute Adherence by Group (Hypothesis 2)

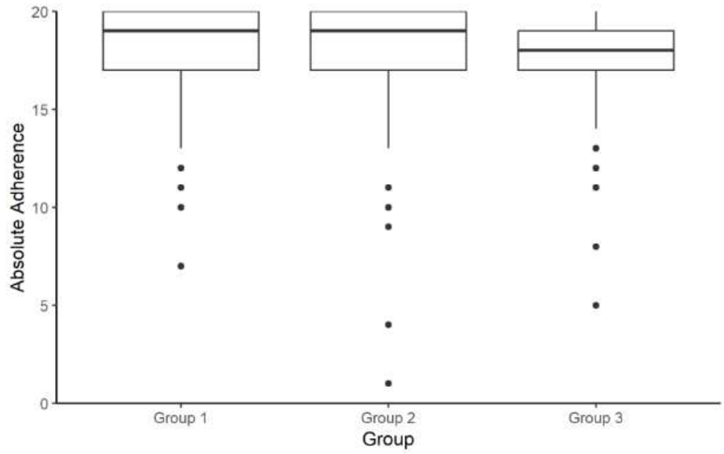
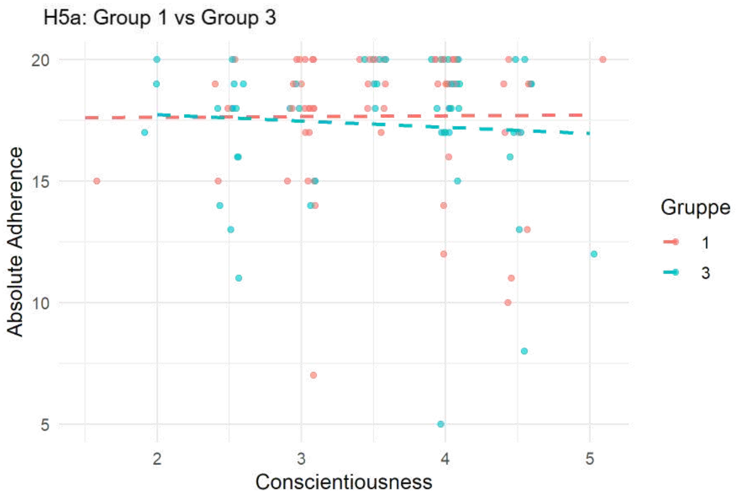
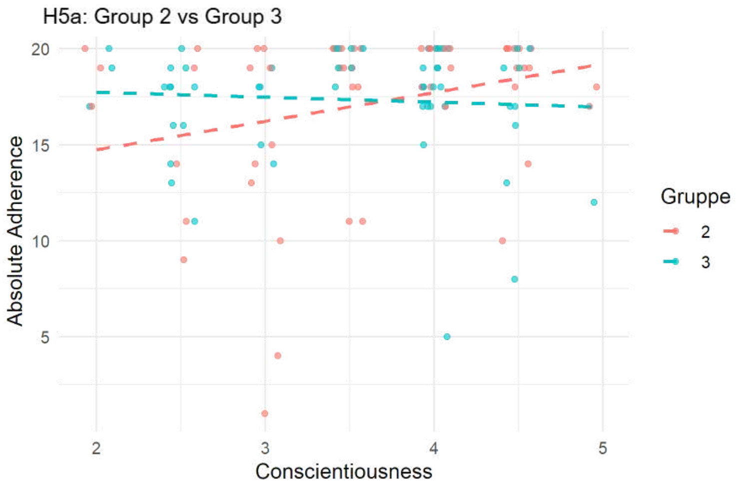


Figure 7

Visual assumption testing for the moderated regression - Scatterplots (Hypothesis 5a)

Scatterplots





Appendix

Appendix A: Recruitment materials (LABS-Text, Flyer, Social Media)

Appendix B: Pre-questionnaire

Appendix C: Emails sent before the online meeting

Appendix D: Form of consent

Appendix E: Examples for tracking schedules

Appendix F: Emails sent after the online meeting

Appendix G: Information about vegetable and fruit consumption

Appendix H: Python code for the schedule creation of participants in group 3

Appendix I: m-Path

Appendix J: Post-questionnaire

Appendix K: Screenshot of G*Power Sample Size Computation

Appendix L: Verwendungserklärung KI

Note: Due to the plagiarism check, the materials from the appendix could not be included.

These can be found on Moodle in the corresponding documentation folder.