To use and engage? Identifying distinct user types in interaction with a smartphone-based intervention

Siezenga, A.M.; Mertens, E C.A.; Gelder, J.-L. van

Computers in Human Behavior Reports 17 (2025) 100602

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

To use and engage? Identifying distinct user types in interaction with a smartphone-based intervention

Aniek M. Siezenga [a,b], Esther C.A. Mertens [a,c], Jean-Louis van Gelder [a,b,*]

[a] Max Planck Institute for the Study of Crime, Security and Law, Günterstalstraße 73, 79100, Freiburg Im Breisgau, Germany
[b] Department of Education and Child Studies, Leiden University, Wassenaarseweg 52, 2333AK, Leiden, the Netherlands
[c] Netherlands Institute for the Study of Crime and Law Enforcement, De Boelelaan 1077, 1081HV, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Smartphone-based intervention
mHealth
Usage data
App engagement
Cluster analysis
User types

ABSTRACT

Background: Smartphone users are a heterogeneous group, implying that certain user types might be distinguishable by the way they interact with a smartphone-based intervention. As these user types potentially benefit differently from an intervention, there is a need to identify them to inform future research, intervention design, and, eventually, improve intervention effectiveness. To this end, we explored 1) whether user types were distinguishable in terms of how much they used a smartphone-based intervention and experienced app engagement; and whether user types differed in 2) their intervention effects; and 3) user characteristics, i.e., HEXACO personality traits and self-efficacy.
Method: Participants were Dutch first-year university students that interacted with the FutureU app aimed at increasing future self-identification. App usage data and engagement survey data were obtained in a randomized controlled trial taking place in 2022 ($n = 86$). *K*-means++ cluster analyses were applied to identify user types based on app use and engagement. Linear discriminant analyses, ANCOVAs, and MANOVAs were conducted to assess whether the clusters differed in intervention outcomes and individual characteristics. The analyses were replicated in data obtained in an RCT taking place in 2023 with an updated version of the app ($n = 106$).
Results: Four user types were identified: Low use–Low engagement, Low use–High engagement, High use–Low engagement, High use–High engagement. Overall, intervention effects were strongest for the user types High engagement–High use and High engagement–Low use. No significant differences were observed in user characteristics.
Conclusion: User types can vary in their use of and engagement with smartphone-based interventions, and benefit differently from these interventions. App engagement appears to play a more significant role than previously assumed, highlighting a need for further studies on drivers of app engagement.

1. To use and engage? Identifying distinct user types in interaction with a smartphone-based intervention aimed at increasing future self-identification

Smartphone users strongly differ in the way they interact with their phones. Some users spend many hours on their phones, others only use them occasionally; some are highly engaged with contents, others less so. These interaction differences also apply to smartphone applications ("apps"[1]) that aim to change the user's behavior – i.e., smartphone-based interventions –, likely resulting in differences in exposure to intervention content. In turn, this might affect intervention effects, necessitating approaches that recognize these individual differences, rather than assuming that all users behave in a similar way (Romano et al., 2022). In this context, identifying distinct user types can help the field move toward understanding why some user types benefit more from smartphone-based interventions than others. Therefore, we aimed to identify distinct user types of a smartphone-based intervention, and examine whether these user types differed in their intervention effects and user characteristics.

* Corresponding author. Max Planck Institute for the Study of Crime, Security and Law, Günterstalstraße 73, 79100, Freiburg im Breisgau, Germany.
E-mail addresses: a.siezenga@csl.mpg.de (A.M. Siezenga), emertens@nscr.nl (E.C.A. Mertens), j.vangelder@csl.mpg.de (J.-L. Gelder).

[1] App: smartphone application; RCT: Randomized Controlled Trial; MARS: Mobile App Rating Scale; TWEETS: TWente Engagement with Ehealth Technologies Scale; FSI: Future Self-Identification; (M)AN(C)OVA: (multivariate) analysis of (co)variance; LDA: Linear Discriminant Analysis, AIC Akaike information criterion, BIC Bayesian information criterion, S_i Silhouette coefficient, CHI Calinski-Harabasz Index.

https://doi.org/10.1016/j.chbr.2025.100602
Received 12 June 2024; Received in revised form 20 December 2024; Accepted 21 January 2025
Available online 24 January 2025

List of abbreviations

App	smartphone application
RCT	Randomized Controlled Trial
MARS	Mobile App Rating Scale
TWEETS	TWente Engagement with Ehealth Technologies Scale
FSI	Future Self-Identification
(M)AN(C)OVA	(multivariate) analysis of (co)variance
LDA	Linear Discriminant Analysis
AIC	Akaike information criterion
BIC	Bayesian information criterion
S_i	Silhouette coefficient
CHI	Calinski-Harabasz Index

1.1. User types

Even though users form a heterogeneous group due to variations in their app use and engagement, prior research has predominantly employed variable-centered approaches, rather than person-centered approaches. The former approach examines relationships between variables, suggesting that all app users behave similarly (e.g., all users use the app frequently and are highly engaged). In contrast, the latter approach identifies subgroups within the data, enabling the differentiation of distinct user types (Muthén & Muthén, 2000) (e.g., a user type characterized by high use frequency and high engagement versus a user type with high use frequency but low engagement). Given that prior studies have identified differences between users, demonstrating that a "standard" app user does not exist (Baumel et al., 2019), it makes sense to identify user types with person-centered approaches.

Previous studies employing person-centered approaches have identified user types based on app usage data. For example, Bell and colleagues (2020) identified three distinct user types in a smartphone-based intervention aimed at alcohol reduction: engagers (kept on using the app), slow disengagers (steadily declined in app use), and fast disengagers (quickly declined in app use). Similarly, Canan and colleagues (2021) identified four user types based on the use of app features in a smartphone-based intervention for clinic patients living with HIV, ranging from frequent to occasional users. These two pioneering studies provided important insight into how user types differ in their *use* of a smartphone-based intervention. However, app interaction entails more than mere use; just as important is engagement.

App engagement entails affective (i.e., enjoyment of progress, and identification with the app and what it stands for), cognitive (e.g., feeling supported in achieving personal goals) and behavioral (i.e., establishing a routine in app use) aspects of the user experience (Kelders et al., 2020). Its significance for interacting with smartphone-based interventions has only recently been highlighted (Perski et al., 2017; Kelders et al., 2020; Siezenga et al., 2023). Furthermore, a recent study recognized app use and app engagement as distinct and uncorrelated constructs (Siezenga et al., 2023). Therefore, studying user types based on app use *and* engagement could help increase our insight into establishing whether user types are distinguishable in terms of both aspects.

1.2. Intervention effects

Identifying user types can help identify who benefits (most) from an intervention and enhances our understanding of individual differences in intervention effects. For instance, if a user type characterized by high levels of app use and engagement benefits more from an intervention compared to a user type low in app use and engagement, this would point toward the importance of app interaction patterns in relation to intervention effects. Furthermore, once user types have been identified, future research can investigate strategies to help users transition to a user type that benefits more, thereby enhancing intervention effectiveness.

Previous research suggested that user types may benefit differently from a smartphone-based intervention (Canan et al., 2021). Specifically, in the study by Canan and colleagues (2021), all user types improved throughout the study period, but the user type that interacted only occasionally with the app had the lowest intervention effects and was least likely to maintain improvement. This study provided important insights into how user types obtain different intervention effects. However, as only app usage data were analyzed, the role of engagement in relation to user types and intervention effects remains unclear. Therefore, the second aim of our study was to explore the relationship between user types and intervention effects.

1.3. User characteristics

Yet the relationship between user types and intervention effects is only one relevant aspect. In this context, it is also of interest to identify user types based on their users' characteristics. This helps in targeting individuals who are most likely to succeed by using a specific intervention, and may enable us to tailor smartphone-based intervention content to fit user needs and preferences. Specifically, we were interested in personality and self-efficacy, as previous research has linked both to treatment outcomes in conventional mental health treatment (Bandura, 1977; Bucher et al., 2019), and to app interaction behavior. To illustrate, previous research studied how personality traits relate to use of smartphone-based interventions. For example, Openness to experience has been positively related to app use (Mikolasek et al., 2018; Su et al., 2020). As the characteristics of people who score high on this trait include curiosity and a preference for experiment (Lee & Ashton, 2004), they might be more open to trying out novel interventions. Moreover, since conscientious individuals are typically thorough and structured (Lee & Ashton, 2004), they may be more inclined to adhere and interact with the app on a daily basis. Consequently, user types characterized by high app use and engagement might be more likely to comprise users with high levels of Openness to experience and Conscientiousness, compared to user types characterized by low app use and engagement.

Additionally, self-efficacy refers to an individual's general belief in their ability to cope with demands and challenges in life (Schwarzer & Jerusalem, 1995). Self-efficacy has been positively related to use of a physical health app (Vinnikova et al., 2020), and it has been proposed that this concept relates positively to app engagement (Perski et al., 2017). Self-efficacy may be relevant in interaction with smartphone-based interventions, as it is a determinant of what activities individuals choose to engage in, the amount of time and effort they are willing to invest, and the persistence they demonstrate when faced with challenges (Bandura, 1977). Therefore, one could expect user types high in app use and engagement to comprise users who score relatively high for self-efficacy compared to user types low in app use and engagement. Both personality traits and self-efficacy remain understudied in the context of user types, but would be constructs that facilitate the identification of users likely to benefit from an intervention. Consequently, the third aim of our study was to establish whether personality characteristics distinguish the identified user types.

1.4. Present study

The present study aimed to identify and explore user types based on both app use and engagement, inspect the relationship between these user types and intervention effects (i.e., future self-identification), and determine relevant user characteristics in this context. We used data from an experimental study in which we studied the effectiveness of the FutureU smartphone-based intervention. This intervention aims to enhance individuals' orientation to the future and to strengthen their identification with who they will be in the future. Our study contributed to the literature in the following ways. First, we identified user types rather than studying the group of users as a whole by applying a person-centered approach. Second, we built on existing research by identifying user types based on both app use *and* engagement. Third, we combined a person-centered approach with a variable-centered approach to analyze differences in intervention effects and user characteristics between the user types. We studied the following research questions:

1) What user types can be distinguished based on their app use and engagement? Although exploratory, we generally assumed that four user groups would emerge based on logical reasoning: One high in both use and engagement, one low in both use and engagement, one high in use, but low in engagement, and one low in use and high in engagement.
2) To what extent do user types differ in terms of intervention effects (i.e., vividness, relatedness, and valence of the future self)? As it is generally assumed that more exposure to and more engagement with a smartphone-based intervention relates to stronger intervention effects, we hypothesized that the user type(s) reflecting both high levels of app use and engagement would relate to stronger intervention effects.3) To what extent do user types differ in user characteristics (i.e., HEXACO personality and self-efficacy)? We approached this research question exploratory.

2. Methods

2.1. Study design

We included data from two randomized controlled trials (RCTs) testing the effectiveness of a smartphone-based intervention aiming to instill a future-oriented mindset, FutureU. Ethical approval was obtained from the independent Ethics Board of the Institute of Education and Child Studies at Leiden University (ECWP2021-320). The RCTs' protocols were prospectively registered in the Netherlands Trial Register (number NL9671) and Clinicaltric-also.gov (number NCT05578755) and published in (Mertens et al., 2022a) and (Mertens et al., 2022b). RCT 1 was conducted from October 2021 to March 2022 and contained a control group and an intervention condition that used the FutureU smartphone-based intervention. The feedback and effectiveness results (Mertens et al., 2024) obtained from this RCT were used to improve the app. To test the updated app, we conducted RCT 2 from October 2022 to July 2023. This second RCT contained a control group and two intervention groups receiving the FutureU intervention either delivered via the smartphone app, or in a Virtual Reality setup. For the present study, we first ran the analyses on data from the intervention condition that used the FutureU app in RCT 1. To test the robustness of the results, we replicated the analyses using data from RCT 2's smartphone app condition.

2.2. Procedure

The procedure for both RCTs was identical. During an intake session, participants gave written informed consent, completed the pre-measurement questionnaire,[2] and installed the FutureU app. Immediately following the 21-day intervention period, participants received the post-measurement questionnaire per e-mail. Participants were considered dropouts if they did not respond to the post-measurement questionnaire. Responses to the post-measurement questionnaire that were not received within 8 days were deleted and treated as missing data. For completing the questionnaires, participants were rewarded with 6 study credits or 25 euro in RCT 1, and 8 study credits or 35 euro in RCT 2.

2.3. Participants

In RCT 1, 87 Dutch first-year university students participated. One participant could not install the app due to a technical error and was excluded from the analyses. Participants in RCT 1 ($n_{\text{female}} = 70$; 81.40%) were on average 19.72 years old ($SD = 3.44$) and mostly studying social sciences ($n = 81$; 94.20%). During RCT 2, 107 Dutch participants were included. Again, one participant could not install the app and was excluded from analyses. Usage data from two participants were deleted because they received the same app installation code, preventing the connection of their app usage data and engagement data. Participants in RCT 2 ($n_{\text{female}} = 97$; 91.51%) were on average 19.40 years old ($SD = 2.46$) and were all studying social sciences ($n = 106$; 100%).

2.4. FutureU app

The FutureU app aims to enhance people's consideration of the future and to strengthen their identification with their older, or "future", selves. This future self-identification involves vividly imagining one's future self, and fostering a sense of relatedness and positive regard towards the future self (Mertens et al., 2022b). The app introduced participants to their future self, represented by a three-dimensional age-progressed avatar based on a photo of the user's face (i.e., selfie), which is made to look roughly 10 years older. The app is built around a scripted daily chat conversation in which participants received psychoeducation, answered questions, completed assignments, and received motivational messages. The 21-day intervention period was divided into three week-long modules. The first module aimed to promote future self-identification; the second module helped participants take the perspective of their future self; and the third module aimed to stimulate a growth mindset (Dweck & Yeager, 2019) and underpin goal achievement.

The daily interactions took approximately 2–5 min. When participants skipped a day or aborted a session, they received the missed content at the next login. Participants in RCT 2 interacted with an updated version of the FutureU app. Although the theoretical framework was similar and the content of both apps overlapped in many ways, certain app components differed. For example, the daily interactions were updated, resulting in different daily app duration patterns. Moreover, the design and usability of features was improved in order to stimulate engagement. For a more extensive overview of module content, see Mertens et al., 2022a and Mertens et la., 2022b.

2.5. Measurements

2.5.1. App use

Usage data registered by the app contained the date, and start and end time of each login. From these data, we calculated the total time in minutes that participants spent in the app ("duration"), the number of times that participants opened the app ("logins"), the total number of days out of the 21-day intervention period that participants interacted with the app ("days"), and the percentage of content exposure[3] ("exposure").

The exposure variable was created by first calculating the daily minimum duration for delivering the messages in the chat and subsequently subtracting this time from the total time that a participant spent

[2] During RCT 1, the Netherlands went into lockdown due to the COVID-19 pandemic (December 19, 2021 to January 26, 2022). There were no significant differences between participants starting before and during lockdown in age, pre-measurement vividness, valence, and relatedness ($F(4,81) = 1.84$, $p = .130$; partial $\eta^2 = .08$), and gender ($\chi^2(1) = 0.73$, $p = .541$).

[3] Since missed intervention content would be shown during the next login, time spent in the app and number of days were not valid measures of exposure to intervention content. That is, if participants spent above average time in the first module, but then stopped using the app, total time spent in the app could be high, even though exposure to content was low.

in the app that day. Negative results indicated that participants did not spend enough time in the app to receive all intervention content for that day. Conversely, positive values signified that participants spent more time in the app than the minimum chat duration, and received the complete day's content. In a second step, the values were categorized as 0 (No exposure, i.e., no login that day), 1 (Spending one to 5 s in the app (i.e., mere opening of the app), 2 (Exposure to some but less than half of the day's intervention content), 3 (Exposure to more than half of the day's content, but less than the complete content), 4 (Full exposure). For example, if the minimum chat duration for Day 3 was 178 s and participants spent 50 s in the app, their Day 3 exposure was categorized as 2, indicating that they were exposed to less than half of the content. Lastly, the sum of these category values was computed over the 21-day intervention period, with a maximum possible score of 84 (category 4 for each of the 21 days). This sum score was then transformed into a percentage, representing a participant's overall exposure to the intervention content (e.g., a sum score of 84 represents 100% exposure to the intervention content).

2.5.2. Engagement

In RCT 1, we assessed generic app engagement with an adapted 3-item version of the Mobile App Rating Scale's section F (MARS; α = .88, Stoyanov et al., 2015). The three items (e.g., "This app is likely to increase my motivation to think about my future") were assessed at post-measurement on 7-point scales ranging from 1 (completely disagree) to 7 (completely agree). In RCT 2, we assessed generic app engagement with the 9-item TWente Engagement with Ehealth Technologies Scale (TWEETS; α = .90, Kelders et al., 2020). The nine items (e.g., "The FutureU app motivates me to think about my future") were measured at post-measurement on 5-point scales ranging from 1 (completely disagree) to 5 (completely agree).

2.5.3. Intervention outcomes

Future Self-Identification (FSI) was assessed at baseline and post-measurement with three subscales: vividness, valence, and relatedness.

Vividness of the future self, i.e., the degree to which people have a vivid image of their future self, was assessed with five items (e.g., "I find it easy to describe myself 10 years from now"; RCT 1 α_{T1} = .94, α_{T2} = .95, RCT 2 α_{T1} = .93, α_{T2} = .92) (Van Gelder et al., 2015) on a 7-point Likert scale ranging from 1 (completely disagree) to 7 (completely agree).

Valence towards the future self, i.e., how positive participants feel about their future self, was measured with one item ("Indicate how you feel when you think about your future") (Hershfield et al., 2009) using a Self-Assessment Manikin (Bradley & Lang, 1994) ranging from 1 (negative) to 9 (positive).

Relatedness to the future self, i.e., how connected and similar participants feel to their future self, was measured with two items (RCT 1 α_{T1} = .51, α_{T2} = .79, RCT 2 α_{T1} = .53, α_{T2} = .82) (Hershfield et al., 2009). Participants rated the extent to which two circles, representing their present and future selves, overlap on a 7-point scale ranging from 1 (no overlap) to 7 (almost complete overlap).

2.5.4. Personality

Personality was assessed at baseline with the 100-item version of the HEXACO-PI-R (i.e., 16 items per trait) in RCT 1, and with the 60-item HEXACO-PI-R (i.e., 10 items per trait) in RCT 2 (e.g., "People often call me a perfectionist"). The HEXACO assesses the following six traits: Honesty-humility ($\alpha_{RCT\ 1}$ = .87, $\alpha_{RCT\ 2}$ = .69), Emotionality ($\alpha_{RCT\ 1}$ = .85, $\alpha_{RCT\ 2}$ = .76), Extraversion ($\alpha_{RCT\ 1}$ = .87, $\alpha_{RCT\ 2}$ = .81), Agreeableness ($\alpha_{RCT\ 1}$ = .87, $\alpha_{RCT\ 2}$ = .77), Conscientiousness ($\alpha_{RCT\ 1}$ = .87, $\alpha_{RCT\ 2}$ = .83), and Openness to experiences ($\alpha_{RCT\ 1}$ = .81, $\alpha_{RCT\ 2}$ = .76). Items were answered on a 5-point Likert scale ranging from 1 (completely disagree) to 5 (completely agree) (Ashton & Lee, 2009; De Vries et al., 2009).

2.5.5. Self-efficacy

Self-efficacy was assessed at baseline with the General Self-efficacy Scale (Schwarzer & Jerusalem, 1995). The scale consists of 10 items (e.g., "I can usually handle whatever comes my way") measured on a 4-point scale ranging from 1 (completely disagree) to 4 (completely agree, $\alpha_{RCT\ 1}$ = .76, $\alpha_{RCT\ 2}$ = .78).

2.6. Statistical analyses

2.6.1. K-means++ cluster analyses

The first research question aimed to identify user types, i.e., "clusters", based on their app use and engagement, and was analyzed with *k*-means++ cluster analyses. This exploratory data analysis technique identifies subgroups (i.e., a pre-defined number of clusters defined as "*k*") in the data for which observations within each cluster are similar, while observations between clusters are different. Clusters are identified in multiple steps. First, *k*-means++ randomly selects *k* predefined center points in the data, i.e., the initial cluster centers. Second, each data point is assigned to its nearest cluster center. Third, the mean of all data points within each initial cluster is calculated, which becomes the updated cluster center. Fourth, this iterative process continues until the data points no longer switch clusters (Arthur & Vassilvitskii, 2007).

Before conducting the *k*-means++ analyses, all variables were standardized. Outliers, which we defined as more than two standard deviations (SD) from the mean, have a substantial impact on *k*-means++ analyses (Nowak-Brzezińska & Gaibei, 2022); consequently, they were winsorized by replacing them with scores at±two SDs from the mean. We conducted four *k*-means++ cluster analyses, each incorporating one of four usage data variables (i.e., duration, logins, days, and exposure) and the app engagement variable (i.e., MARS or TWEETS). We could not combine multiple usage variables in a single model due to multicollinearity. For each combination of the usage data variable and app engagement (i.e., model), we explored the cluster solutions ranging from 2 to 5 pre-defined clusters, using SPSS 28, resulting in 16 *k*-means++ cluster analyses. To determine the optimal number of clusters, we assessed each model on 1) cluster size (i.e., each cluster containing at least 3% of the total cases), 2) theoretical relevance and interpretability of the clusters, and 3) model fit indices including AIC, BIC,[4] Calinski-Harabasz (CH) Index, and Silhouette coefficient (S_i). Lower AIC and BIC values, a higher CH index, and an S_i value closer to 1 indicate superior model fit. Subsequently, we studied whether the cluster indication variables (i.e., the usage data variable and app engagement) within each *k*-means++ solution differed significantly from one another, using Analyses of Variance (ANOVAs) with Bonferroni-corrected post-hoc analyses.

2.6.2. Linear discriminant analyses

The second and third research questions were investigated by conducting linear discriminant analyses (LDAs). LDA aims to model maximized separation between existing clusters and minimized within-cluster separation by calculating linear functions of a-priori selected predictor variables and identifying which predictor variables are most important in distinguishing between clusters (Vannatta & LaVenia, 2020). For the second research question, we aimed to study whether differences in the intervention effects (i.e., vividness, valence, and relatedness) could differentiate the identified clusters. We reduced multicollinearity risk by conducting three separate LDAs per model, including either post-measurement vividness, valence, or relatedness, while controlling for their pre-measurement values. This resulted in 12 LDAs. For the third research question, we attempted to study whether differences in personality traits and self-efficacy could differentiate

[4] We pre-registered to compare the clusters on the *F*-value and iteration history as well, but as the values were relatively similar and, therefore, inconclusive, we decided to include AIC and BIC values instead.

between the clusters. We ran one LDA per model in which the linear function was based on the six personality traits and self-efficacy.

To assess whether the linear functions were able to significantly distinguish among the identified clusters, Chi[2] was provided and interpreted. In case of significant differences between the clusters (i.e., $p <$ .05), we determined which variable (i.e., research question 2: baseline or post-measurement vividness, valence, or relatedness; research question 3: the personality trait and/or self-efficacy) had the highest correlation with the discriminant function, and was therefore most important for differentiating between the clusters. Fit indicators are provided in Table A.1 in Appendix A.

Subsequently, we ran an additional series of ANCOVAs (for research question 2) and multivariate ANOVAs (MANOVAs) (for research question 3) to assess comparability of the LDA findings and to obtain Bonferroni-corrected post-hoc tests to further analyze potential differences between the clusters. For research question 2, the intervention outcomes vividness, valence, and relatedness at post-measurement were included as independent variables; the clusters as the dependent variables; and baseline vividness, valence, and relatedness as covariates. The HEXACO traits and self-efficacy were included as independent variables for research question 3; the clusters were the dependent variable. Significance was based on $p < .05$, and η^2 effect sizes were interpreted (i.e., $\eta^2 < .06$ = small effect, $.06 > \eta^2 < .4$ = moderate effect, $\eta^2 > .14$ = large effect) to ascertain whether the findings were meaningful.

3. Results

3.1. Study attrition and missing data

Attrition analyses showed no significant differences between study dropouts ($n_{\text{RCT 1}} = 1$ (1.16%); $n_{\text{RCT 2}} = 4$ (3.77%)) and completers in forms of age, gender, education, and outcomes at baseline (see Appendix B.1). In both RCTs, the percentage of missing data was low (RCT 1 = 4.65%; RCT 2 = 4.70%). We handled missing data with listwise deletion, as Little's MCAR test showed a random pattern of missing data using a significance cut-off value of .010 (Van Ness et al., 2007), ($\chi^2_{\text{RCT }1}(14) = 11.65$, $p = .634$; $\chi^2_{RCT\ 2}(32) = 53.13$, $p = .011$).

3.2. Descriptive engagement and app use statistics

App use statistics (see Table 1) showed that, on average, participants were exposed to over 80% of the intervention content ($M_{RCT\ 1} = 83.60\%$; $M_{RCT\ 2} = 87.23\%$). Usage data from both RCTs revealed a general decline in use over the course of the three intervention weeks. From post-measurement engagement scores, it could be concluded that participants, on average, felt engaged with the FutureU app (see Table 1).

Participants who did not log in during the third module ($n_{RCT\ 1} = 8$, $n_{RCT\ 2} = 8$) did not differ from those who did in terms of individual characteristics (i.e., age, gender, and education) and outcomes at baseline (i.e., vividness, relatedness, and valence) (see Appendix B.2 for the statistics).

3.3. K-means++ analyses

Based on the number of cases per cluster, theoretical relevance and interpretability, and model fit indices (see Table 2), the 4-cluster solutions generally fitted the data best in both RCT 1 and RCT 2. These four clusters can be described as: 1) below average use and engagement: "Low use–Low engagement"; 2) below average use and above average engagement: "Low use–High engagement"; 3) above average use and below average engagement: "High use–Low engagement"; and 4) above average use and above average engagement: "High use–High engagement" (see Fig. 1 for a graphical representation of the clusters based on duration, see Appendix C for the graphical representations of the clusters based on logins, days, and exposure).

Despite the 4-cluster pattern in the data, the fit statistics did not unambiguously indicate this pattern for each model (see Table 2). To illustrate, for the model of MARS and Duration (RCT 1), three of the four fit statistics suggested a 5-cluster solution. The fifth cluster was a group of participants representing an average scoring group (i.e., a *z*-score of 0 for both duration and engagement). The other four clusters that still followed the identified 4-cluster pattern were more extreme in their high and low scores. As the fifth cluster was theoretically not of interest, the more parsimonious 4-cluster solution was favored. Furthermore, the 3-cluster solution that was suggested for the model of TWEETS and logins (RCT 2) lacked the theoretically relevant cluster High use–High engagement. As this cluster appeared in the 4-cluster solution, we opted for the 4-cluster solution over the 3-cluster solution based on theoretical justification.[5]

Next, we conducted ANOVAs with Bonferroni-corrected post-hoc tests to compare the four identified clusters with regard to app usage data and engagement. In both RCTs, most clusters differed significantly on the usage data variables. Furthermore, the difference between the two clusters with above average engagement and the two clusters with below average engagement was significant. This indicates that the cluster analyses identified distinct groups of user types (see Appendix D, Tables D.1–D.2 for the Bonferroni-corrected post-hoc results).

Table 1
Descriptive statistics of app use and engagement for the complete sample and per user type.

RCT 1		C1 L/L (*n* = 31)	C2 L/H (*n* = 14)	C3 H/L (*n* = 9)	C4 H/H (*n* = 28)
	M(SD)	*M(SD)*	*M(SD)*	*M(SD)*	*M(SD)*
Usage data	*n* = 86				
Duration min	61.99 (27.94)	57.46 (12.43)	28.57 (13.53)	110.87 (39.27)	69.20 (15.24)
Logins	28.63 (10.82)	29.23 (8.92)	19.14 (13.42)	36.78 (11.40)	29.75 (7.33)
Day	14.53 (4.74)	15.23 (3.86)	9.00(4.98)	18.11 (3.02)	15.71 (3.91)
Exposure %	83.60 (18.77)	88.94 (10.35)	55.95 (22.95)	96.56 (4.01)	88.61 (12.25)
Engagement	*n* = 82				
MARS	4.27 (1.52)	2.92 (1.12)	4.62(0.99)	3.85(1.08)	5.73(0.59)
RCT 2		C1 L/L (*n* = 9)	C2 L/H (*n* = 32)	C3 H/L (*n* = 23)	C4 H/H (*n* = 35)
	M(SD)	*M(SD)*	*M(SD)*	*M(SD)*	*M(SD)*
Usage data	*n* = 104				
Duration min	76.23 (25.65)	31.20 (12.94)	65.13 (10.24)	80.66 (13.09)	99.81 (15.33)
Logins	32.78 (13.76)	14.22 (6.61)	30.19 (13.20)	37.70 (13.01)	37.94 (10.51)
Days	14.63 (5.13)	5.44 (1.81)	13.50 (4.24)	15.83 (3.23)	17.91 (3.33)
Exposure %	87.23 (19.52)	46.56 (20.96)	88.28 (10.34)	93.27 (6.61)	97.89 (4.02)
Engagement	*n* = 102				
TWEETS	3.56 (0.72)	2.62 (0.71)	3.84(0.27)	2.75(0.51)	4.00(0.39)

Note. C1 = Low use–Low engagement, C2 = Low use–High engagement, C3 = High use–Low engagement, C4 = High use–High engagement; MARS = Mobile App Rating Scale, TWEETS = TWente Engagement with Ehealth Technologies Scale. The reported statistics are based on the model that included duration and engagement (i.e., MARS or TWEETS), as the statistics were highly similar across models (see Appendix C for all results).

[5] To assess the robustness of the results, the *k*-means cluster analyses were rerun twice. First, without winsorizing the scores; second, excluding 10 participants in RCT 1 and 14 participants in RCT 2 who experienced technical problems with the app (e.g., the future self avatar could not be created). Generally, the results were roughly similar to those of the main analyses; though there was more ambiguity in the fit statistics (see Appendix E, Tables E.1-E.4).

Table 2
The fit information for the 4 models of RCT 1 (n = 86) and RCT 2 (n = 99).

k-means++ cluster analysis						
			AIC	BIC	CHI	S_i
RCT 1		*k*				
	Duration	2	63.91	83.17	54.08	.518
		3	52.56	81.44	53.67	.507
		4	49.64	88.15	49.22	.495
		5	45.11	93.24	63.55	.535
	Logins	2	78.49	97.74	49.94	.494
		3	66.13	95.01	56.47	.501
		4	58.05	96.56	68.40	.532
		5	60.27	108.40	62.60	.534
	Day	2	87.27	106.52	41.61	.465
		3	70.10	98.98	59.26	.509
		4	66.33	104.84	62.17	.522
		5	68.61	116.75	55.04	.468
	Exposure	2	56.77	76.02	59.44	.524
		3	36.48	65.36	76.06	.560
		4	35.37	73.88	71.71	.555
		5	37.41	85.54	68.48	.536
RCT 2	Duration	2	64.46	85.23	71.31	.542
		3	50.11	81.25	67.59	.488
		4	41.42	82.94	79.17	.549
		5	41.30	93.21	76.57	.520
	Logins	2	82.87	103.62	56.63	.479
		3	58.38	89.53	78.15	.533
		4	56.69	98.21	69.49	.504
		5	54.30	106.20	71.20	.490
	Day	2	75.23	95.99	66.84	.519
		3	59.05	90.19	72.05	.523
		4	46.74	88.26	90.70	.566
		5	49.74	101.64	80.19	.547
	Exposure	2	7.77	28.53	95.48	.619
		3	−9.35	21.79	104.69	.629
		4	−15.20	26.32	99.27	.599
		5	−16.34	35.56	113.70	.565

AIC Akaike information criterion, *BIC* Bayesian information criterion, *CHI* Calinski-Harabasz Index, S_i Silhouette coefficient.

4. Intervention effects

4.1. Linear discriminant analyses based on intervention effects

We ran LDAs to study whether the intervention baseline and post-measurement variables vividness, relatedness, and valence were able to distinguish the clusters. The results gave indications for a relationship between intervention effects and cluster membership, as four out of 12 models were significant in RCT 1, and 12 out of 12 models in RCT 2.

In RCT 1, both functions based on baseline and post-measurement valence were together able to distinguish between one or more clusters in the model of duration ($Chi^2(6) = 13.97$, $p = .030$). In the model of exposure, both functions based on baseline and post-measurement vividness were together able to do so ($Chi^2(6) = 13.08$, $p = .042$). In the model of logins, this was the case for both functions based on baseline and post-measurement vividness ($Chi^2(6) = 20.00$, $p = .003$) and baseline and post-measurement relatedness ($Chi^2(6) = 18.05$, $p = .006$). In RCT 2, in all four models, both functions based on baseline and post-measurement vividness, baseline and post-measurement relatedness, and baseline and post-measurement valence were together able to distinguish between one or more clusters ($Chi^2(6) = 21.83–19.84$, $p = <.001–.010$). In all significant discriminant functions, the post-measurement variable was the highest contributor to the discriminant function, indicating that the post-measurement intervention outcome was most important in distinguishing the clusters. An extended overview of the LDA fit indicators can be found in Appendix A, Tables A.1–A.2.

4.2. ANCOVAs

We carried out ANCOVAs to compare the four clusters on their intervention effects. Findings from these analyses were in line with the LDA results, indicating that, controlling for baseline, the clusters differed in intervention effects. RCT 1 data showed differences between the clusters, as in six out of 12 models there was a significant difference between at least two clusters. More specifically, with moderate to large effect sizes ($\eta^2 = .06–.14$), we identified differences between the clusters three times in vividness, two times in relatedness, and one time in valence (see Table 3). Results from RCT 2 data were more pronounced, as in 12 out of 12 models we identified significant differences between at least two clusters in vividness, relatedness, and valence, with moderate to large effect sizes ($\eta^2 = .10–.19$). In both RCT 1 and RCT 2 data, the ANCOVAs showed that the High use–High engagement cluster generally increased more on vividness, relatedness, and valence than the Low use–Low engagement cluster. In RCT 2 data, there was an additional pattern suggesting that both the High use–High engagement cluster and the Low use–High engagement cluster generally increased more on vividness, relatedness, and valence compared to the Low use–Low engagement cluster and the High use–Low engagement cluster (see Table 3).

5. User characteristics

In order to identify user types based on user characteristics, we also

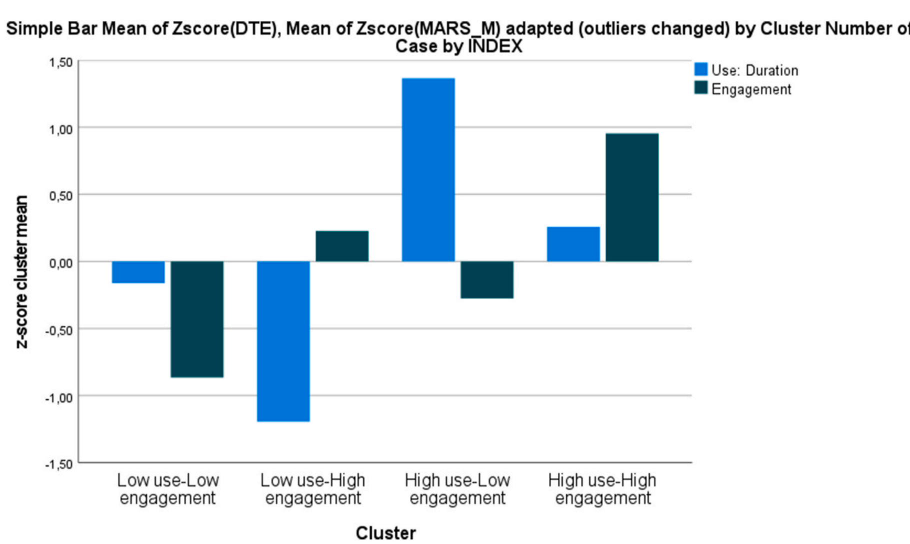

Fig. 1. Graphical representation of the cluster formation based on duration and engagement in RCT 1.

Table 3
ANCOVAs of the relationship between clusters and intervention effects.

		RCT 1 (n = 82)							RCT 2 (n = 99)				
					ANCOVA						ANCOVA		
Model		Clusters	T1 M (SD)	T4 M (SD)	F (df, df)	p	$\eta^2 partial$	Clusters	T1 M (SD)	T4 M (SD)	F (df, df)	p	$\eta^2 partial$
Duration													
	Vividness	C1	3.17 (1.45)	3.43 (1.44)	2.08 (3,77)	.110	.08	C1	2.67 (1.26)	3.82 (1.12)	7.36 (3,94)	<.001	.19
		C2	3.37 (1.18)	3.81 (1.33)				$C2_{a}$	3.51 (1.57)	4.54 (1.13)			
		C3	2.47 (0.44)	3.44 (1.94)				$C3_{ab}$	3.07 (1.40)	3.44 (1.43)			
		C4	3.64 (1.50)	4.39 (1.22)				$C4_{b}$	3.27 (1.36)	4.78 (0.98)			
	Relatedness	C1	3.62 (1.28)	3.98 (1.13)	0.79 (3,77)	.505	.03	$C1_{a}$	3.67 (1.06)	3.56 (1.04)	5.04 (3,94)	.003	.14
		C2	3.89 (1.35)	4.00 (1.30)				C2	3.67 (1.04)	4.30 (1.00)			
		C3	3.72 (0.44)	4.11 (0.86)				$C3^{c}$	4.09 (0.93)	4.37 (1.14)			
		C4	3.93 (0.86)	4.46 (0.91)				$C4^{c}_{a}$	3.77 (0.98)	4.76 (1.00)			
	Valence	C1	6.26 (1.84)	6.42 (1.23)	3.99 (3,77)	.011	.14	C1a	6.00 (2.06)	6.00 (1.73)	5.68 (3,94)	.001	.15
		C2a	6.71 (1.49)	5.86 (1.66)				C2	6.69 (1.23)	7.13 (0.87)			
		C3	6.89 (1.90)	6.00 (1.50)				$C3_{b}$	6.61 (1.56)	6.52 (1.47)			
		$C4_{a}$	6.93 (1.12)	6.93 (1.15)				$C4_{ab}$	6.69 (1.39)	7.40 (0.95)			
Logins													
	Vividness	$C1^{c}_{a}$	2.68 (1.57)	2.73 (1.32)	4.83 (3,77)	.004	.11	$C1^{c}$	2.97 (1.30)	3.67 (1.21)	4.01 (3,94)	.010	.11
		C2	3.06 (1.20)	3.67 (1.29)				$C2^{cd}$	3.19 (1.42)	4.52 (1.05)			
		$C3^{c}$	3.63 (1.27)	4.24 (1.30)				$C3^{d}$	2.82 (1.56)	3.40 (1.71)			
		$C4_{a}$	3.77 (1.71)	4.55 (1.32)				C4	3.63 (1.47)	4.69 (1.16)			
	Relatedness	$C1^{c}_{a}$	3.14 (1.11)	3.33 (0.80)	3.28 (3,77)	.025	.01	$C1^{c}$	3.92 (0.91)	4.06 (1.16)	3.29 (3,94)	.024	.10
		$C2_{a}$	4.08 (1.05)	4.54 (0.91)				C2	3.51 (0.96)	4.34 (1.06)			
		C3	3.97 (1.26)	4.23 (1.35)				C3	4.10 (0.99)	4.15 (1.11)			
		$C4^{c}$	3.83 (0.84)	4.35 (0.91)				C4c	4.03 (0.97)	4.80 (0.96)			
	Valence	C1	5.83 (1.79)	5.83 (1.54)	0.30 (3,77)	.827	.11	$C1_{ab}$	6.39 (1.82)	6.28 (1.53)	5.12 (3,94)	.003	.14
		C2	6.80 (1.32)	6.56 (1.36)				$C2^{c}_{a}$	6.78 (1.35)	7.29 (0.87)			
		C3	6.73 (1.91)	6.53 (0.83)				$C3^{cd}$	6.60 (1.35)	6.40 (1.43)			
		C4	7.00 (1.29)	6.75 (1.39)				$C4^{d}_{b}$	6.50 (1.38)	7.17 (1.18)			
Day													
	Vividness	$C1^{cd}$	3.15 (1.54)	3.13 (1.59)	3.01 (3,77)	.035	.11	C1a	2.73 (1.48)	3.60 (1.27)	5.89 (3,94)	<.001	.16
		$C2_{c}$	3.44 (1.50)	4.27 (1.18)				C2c	3.42 (1.52)	4.50 (0.95)			
		C3	3.13 (1.51)	3.61 (1.46)				$C3^{c}_{b}$	2.98 (1.23)	3.53 (1.43)			
		$C4^{d}$	3.37 (1.52)	4.07 (1.41)				$C4_{ab}$	3.44 (1.44)	4.79 (1.10)			
	Relatedness $C1^{c}$	3.38 (1.26)	3.59 (1.13)	2.97 (3,77)	.037	.10	$C1^{c}_{a}$	3.86 (1.05)	3.79 (1.12)	4.80 (3,94)	.004	.13	
		C2	3.68 (0.80)	4.14 (1.07)				$C2^{c}$	3.67 (1.12)	4.38 (1.00)			
		C3	3.79 (1.29)	3.85 (1.26)				C3	3.95 (0.83)	4.29 (1.12)			
		$C4^{c}$	4.00 (0.99)	4.59 (0.92)				$C4_{a}$	3.80 (0.95)	4.70 (1.02)			
	Valence	C1	5.94 (1.95)	5.94 (1.53)	0.38 (3,77)	.770	.01	$C1^{c}_{a}$	6.29 (1.82)	6.21 (1.67)	5.26 (3,94)	.002	.14

(continued on next page)

Table 3 (*continued*)

		RCT 1 (n = 82)						RCT 2 (n = 99)					
					ANCOVA						ANCOVA		
Model		Clusters	T1 *M* (*SD*)	T4 *M* (SD)	*F* (*df, df*)	*p*	η^2*partial*	Clusters	T1 *M* (*SD*)	T4 *M* (*SD*)	*F* (*df, df*)	*p*	η^2*partial*
		C2	6.71 (1.07)	6.29 (1.49)				$C2^{cd}$	6.50 (1.33)	7.12 (1.03)			
		C3	6.41 (1.91)	6.47 (1.18)				$C3^{d}_{b}$	6.68 (1.63)	6.58 (1.46)			
		C4	7.03 (1.29)	6.74 (1.27)				$C4_{ab}$	6.75 (1.30)	7.35 (0.83)			
Exposure													
	Vividness	C1	3.10 (1.79)	3.20 (1.98)	3.14 (3,77)	.030	.06	C1	2.53 (1.37)	3.62 (1.37)	6.77 (3,94)	<.001	.18
		C2	3.15 (1.32)	4.07 (1.25)				C2	3.58 (1.96)	4.45 (1.29)			
		$C3_{c}$	2.96 (1.35)	3.31 (1.31)				$C3_{a}$	3.11 (1.31)	3.54 (1.34)			
		$C4^{c}$	3.74 (1.59)	4.45 (1.31)				$C4_{a}$	3.34 (1.34)	4.73 (1.00)			
	Relatedness	C1	3.44 (1.29)	3.56 (1.29)	1.66 (3,77)	.183	.05	$C1_{a}$	3.78 (1.03)	3.50 (1.06)	5.65 (3,94)	.001	.15
		C2	3.79 (0.99)	4.17 (0.78)				C2	3.62 (1.06)	4.15 (0.92)			
		C3	3.59 (1.19)	3.92 (1.15)				C3	3.98 (0.98)	4.33 (1.13)			
		C4	4.07 (0.93)	4.58 (0.89)				$C4_{a}$	3.77 (0.96)	4.66 (1.01)			
	Valence	C1	6.00 (2.33)	5.63 (1.85)	1.23 (3,77)	.304	.11	$C1_{a}$	6.22 (1.79)	5.78 (1.86)	6.48 (3,94)	<.001	.17
		C2	7.00 (1.41)	6.25 (1.66)				C2	6.92 (1.04)	7.08 (0.76)			
		C3	6.34 (1.58)	6.41 (1.01)				$C3_{b}$	6.42 (1.74)	6.58 (1.35)			
		C4	6.97 (1.35)	6.80 (1.35)				$C4_{ab}$	6.68 (1.33)	7.34 (0.94)			

Note. C1 = Low use–Low engagement, C2 = Low use–High engagement, C3 = High use–Low engagement, C4 = High use–High engagement, T1 = pre-measurement, T4 = post-measurement.
Clusters sharing subscript “a” and “b” differ significantly from each other ($p < .05$), and clusters sharing superscript “c” and “d” show a trend towards significant differences from another ($p < .10$) as indicated by Bonferroni.

ran LDAs to study whether personality traits and self-efficacy were able to distinguish between the clusters. In general, the majority of the LDAs in both RCTs resulted in non-significant discriminant functions, indicating that personality traits and self-efficacy were not able to significantly distinguish the four clusters (see Appendix F). This pattern was confirmed by the MANOVAs, generally revealing no differences across clusters in personality and self-efficacy (see Table 4 for main effects).

In line with the LDA findings, post-hoc analyses revealed that in the model of MARS and Day (RCT 1), the clusters differed significantly in Conscientiousness ($F(3,78) = 3.94$, $p = .011$, $\eta^2 = .13$), indicating a moderate to large effect. Furthermore, in the model for TWEETS and logins (RCT 2), the clusters differed significantly in Openness to experience levels ($F(3,95) = 4.51$, $p = .005$, $\eta^2 = .13$), indicating a moderate to large effect. Tables F.1-F.2 in Appendix F provide an overview of all post-hoc tests results.

6. Discussion

Analyzing user types of smartphone-based interventions can help the field move toward a better understanding why some user types derive greater benefits from smartphone-based interventions than others. To the best of our knowledge, the present study is the first to study user types within a smartphone-based intervention based on both app use *and* engagement. In a sample consisting of Dutch first-year university students interacting with an app intervention aimed at instilling future self-identification, we examined what user types can be distinguished from one another, whether these user types benefitted differently from the intervention, and whether user types could be identified based on differences in personality traits and self-efficacy.

As we identified four distinct user types characterized by different levels of app use and engagement – Low use–Low engagement, Low use–High engagement, High use–Low engagement, High use–High engagement – we confirmed that users of smartphone-based interventions cannot be seen as a homogenous group. Notably, the user

Table 4
Main effects of MANOVA regarding differences between clusters in personality and self-efficacy.

MANOVA								
RCT 1					RCT 2			
Model	Wilks’ Lambda	*F (df, df)*	*p*	η^2	Wilks’ Lambda	*F (df, df)*	*p*	η^2
Duration	.654	1.57 (21, 207.30)	.058	.13	0.71	1.58 (21, 256,11)	.055	.11
Logins	.740	1.09 (21, 207.30)	.360	.10	0.68	1.74 (21, 256,11)	.025	.12
Day	.629	1.73 (21, 207.30)	.029	.14	0.75	1.27 (21, 256,11)	.198	.09
Exposure	.729	1.15 (21, 207.30)	.299	.10	0.79	1.04 (21, 256,11)	.420	.08

types responded differently to the intervention, with the user type high in both app use and engagement benefiting most from the intervention content. In addition, there were indications that the user types high in engagement (i.e., Low use–High engagement and High use–High engagement) derived greater intervention benefits compared to the user types low in engagement (i.e., High use–Low engagement and Low use–Low engagement). Differences in user type did not seem attributable to personality or self-efficacy. The similarity in findings across two RCTs underpins the robustness of our findings.

6.1. User types

Our classification into four distinct user types adds to prior work by revealing variability not only in app use (e.g., Bell et al., 2020; Canan et al., 2021), but also in level of engagement. Identifying user types characterized by low use but high engagement, and high use but low engagement challenges current theoretical models of app engagement that suggest that high app use implies high engagement, and vice versa (e.g., Perski et al., 2017; Yardley et al., 2015). This finding may seem surprising, but is in line with prior research emphasizing that app use and engagement are distinct constructs that are both important aspects of app interaction (Siezenga et al., 2023; Kelders, 2015; Kelders et al., 2020). After all, Low use–High engagement users could have had time constraints or already internalized behavior change strategies to the extent that they no longer depended on the app to achieve their goals (Yardley et al., 2015). Users in the High use–Low engagement user type category might have persisted in using the app when the perceived benefits outweighed the costs of participating (Donkin & Glozier, 2012).

Finding different user types of a smartphone-based intervention has implications for research. First, researchers are advised to more often consider the use of statistical approaches that capture heterogeneity of users (e.g., person-centered approaches). Second, considering the observed differences in user types' app engagement, future research should include measures to capture app engagement. Third, current app engagement models in which app use is seen as part of app engagement should be revised. Fourth, it is important to design smartphone-based interventions that fit diverse user types, which may be achieved through tailoring content and enabling personalization of apps (Hollis et al., 2017; Yardley et al., 2015).

6.2. Intervention effects

Our findings also revealed that user types benefitted differently from an intervention, with those high in app use and engagement benefitting most. Perski and colleagues (2017) have suggested that higher app use and engagement increase motivation and skills for behavior change, which relate to stronger intervention effects. Moreover, as both user types high in engagement seemed to benefit more from the FutureU intervention compared to the user types low in engagement, engagement might be a more critical factor than use in relation to intervention effects. Even though this finding did not appear in the data of the first RCT and should therefore be interpreted with caution, it is in line with a study by Kelders (2015), who studied predictors of intervention effects of an internet-based intervention. In this study, involvement, i.e., an attribute of cognitive engagement, outperformed the length of intervention use as a predictor of intervention effects (Kelders, 2015). As engagement, and not use, has been suggested to share attributes with focused attention, feeling engaged potentially enhances users' receptivity to information and behavior change more than mere app use (O'Brien & Toms, 2009).

The finding that both user types high in engagement experienced the greatest benefits from the intervention content suggests the necessity of transitioning users from low to high engagement user types. That said, accomplishing this transition presents a challenge due to the scarcity of research addressing whether users themselves or the app need to be targeted to support this shift. That is, app engagement is commonly approached either as a user characteristic, with social scientists investigating the degree to which users experience engagement, or as an app characteristic, with human-computer interaction researchers analyzing what app features or intervention-related factors elicit engagement. Approaching engagement as a user characteristic suggests that certain users may be more or less likely to feel engaged with smartphone-based interventions. This "trait-like" perspective on engagement implies that such interventions may not be suitable for all users. When engagement is approached as an app characteristic, strategies for enhancing engagement are looked for in the technology. This can be achieved, for example, through integrating engaging features like gamification, or designing interventions that adapt to user needs through tailored content or personalized app interfaces. While both approaches provide valuable insights, interdisciplinary research on the interplay between engagement as a user characteristic, engagement as an app characteristic, and intervention effects is needed. Such studies could increase our understanding of the dynamics between user and app characteristics in relation to intervention effects, and inform strategies aimed at transitioning users to highly engaged user types.

6.3. User characteristics

While much of the required research on the implications of user characteristics remains to be done, some factors were covered by the present study. Regarding user characteristics, user types could not be identified based on users' personality traits and self-efficacy. This may appear contradictory to prior research indicating a relationship between personality traits and app use (e.g., Beierle et al., 2020). However, our study is different from previous research in two ways: First, we examined the relationship between user characteristics and user types, rather than treating users as a homogenous group. Second, classification of user types not only relied on app use, but also incorporated measures of engagement. Preliminary evidence suggests a lack of association between personality traits and app engagement (Kelders et al., 2020; Khwaja et al., 2021). For example, Khwaja and colleagues (2021) observed that correlations between personality traits and smartphone-based intervention use (i.e., number of completed activities per intervention component) did not overlap with correlations between personality traits and user app engagement (i.e., positive ratings of intervention components). These findings suggest that app use and engagement are driven by different processes. According to Kelders and colleagues (2020), engagement should reflect more intrinsic motivating reasons for interacting with an app intervention, rather than being dictated by someone's personality profile (Kelders et al., 2020). Consequently, the assumed association between user types and user characteristics such as personality traits and self-efficacy might be incorrect.

Nevertheless, future research should continue to explore predictors of user types. Previous research identified significant associations between sociodemographic characteristics (age, gender, and educational qualifications) and behavioral engagement (Garnett et al., 2018). Therefore, future research involving more socio-demographically diverse samples could examine whether these characteristics cane effectively distinguish user types. However, rather than focusing solely on user traits, it may be more relevant to explore factors closely related to app interaction, such as users' motivation for and expectations of smartphone-based interventions (Perski et al., 2017). For example, users vary in their motivation to download a smartphone-based intervention and in their expectations of the app's efficacy (Romano et al., 2022). Users exhibiting high motivation for interacting with such interventions may be more likely to end up in a highly engaged user type, while those with low expectations may display reduced motivation and gravitate towards low engagement user types.

6.4. Limitations

The first study limitation is its incentivized nature. Even though participants received compensation for filling in the questionnaires, and not for using the app, their motivation to interact with the app might nevertheless have been affected, as they might have felt a research-related obligation to adhere (Donkin & Glozier, 2012). Second, the generalizability of our findings may be limited due to the use of a nearly identical smartphone-based intervention in both RCTs and the homogenous samples in terms of socio-demographic characteristics. The specific aims, goals, and features of the FutureU app may have let to app-specific interaction patterns that do not generalize to other app interventions lacking these elements. Therefore, we encourage future studies to examine whether the findings reported here extend to other smartphone-based interventions and samples. Third, the sample size for some identified user types was relatively small. Although our smallest subgroup comprised 10% of the total sample, which exceeds the acceptable minimum (Galatzer-Levy et al., 2013), we encourage future researchers to apply cluster analyses to larger samples. Smartphones facilitate such large-scale data collection, enabling replication of our study in other app interventions and more heterogenous samples. Furthermore, due to the high multicollinearity between the vividness, valence, and relatedness variables, we had to run separate analyses instead of including them in the same model, which increased the risk of an inflated alpha in the LDAs, ANCOVAs, and MANOVAs. Therefore, instead of interpreting each individual result, we interpreted the patterns and analyzed effect sizes to gain insight into the strength of the results. In addition, the robustness of the results was checked by running the analyses in two different datasets. Lastly, the reliability of the relatedness variable at baseline was low. Although reliability improved in subsequent measurements and the results follow patterns similar to those of presence and valence, we advise caution in interpreting the specific findings involving baseline relatedness.

7. Conclusion

First-year university students interacting with a smartphone-based intervention aimed at instilling future self-identification cannot be considered a homogeneous group, as we identified four user types based on their use of, and engagement with, a smartphone-based intervention. Where we found differences between the user types in the extent to which they benefitted from the intervention content, we did not find indications for differences between the user types in forms of personality and self-efficacy. Our findings emphasize the importance of app engagement in relation to intervention effects. Specifically, there is a need for experimental studies in which engaging app features (e.g., gamified elements) are manipulated to study how app characteristics, user characteristics, and intervention outcomes are related. Given that our results suggest that there are distinct user types and that these relate differently to intervention effects, it seems important to design smartphone-based interventions that are able to cater to diverse user types (e.g., through tailoring content and personalization) in order to increase the potential of smartphone-based interventions for all.

CRediT authorship contribution statement

Aniek M. Siezenga: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Esther C.A. Mertens:** Writing – review & editing, Supervision, Resources, Methodology, Investigation, Conceptualization. **Jean-Louis van Gelder:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

Ethics approval and consent to participate

The study was performed in line with the principles of the Declaration of Helsinki. Ethical approval was obtained by the Ethics Board of the Institute of Education and Child Studies at Leiden University (ECWP2021-320) on March 21, 2021. All participants gave written informed consent to participate in the study.

Consent for publication

We hereby declare that we have obtained written informed consent from all subjects for publication of identifying information/images.

Availability of data and materials

The dataset analyzed during the current study will be available after publication in the Dataverse repository (https://doi.org/10.34894/C3BKJ2).

Funding

This work was supported by the ERC Consolidator Grant [Grant Number 772911-CRIMETIME, 2018]. The funder was not involved in the study design, data collection, analyses, data interpretation and writing of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for *Computers in Human Behavior Reports* and was not involved in the editorial review or the decision to publish this article.

Acknowledgements

The authors would like to thank Orb Amsterdam (http://www.orbamsterdam.com) for developing the FutureU application, Dr. R. Rippe for the statistical consultation, and Esther Groot and Jip Otte for their assistance in data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbr.2025.100602.

Data availability

Data will be made available on request.

References

Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. In *In proc annu ACM-SIAM symp discrete algorithms* (pp. 1027–1035). SODA '07), USA.

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*(4), 340–345. https://doi.org/10.1080/00223890902935878

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.

Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective user engagement with mental health apps: Systematic search and panel-based usage analysis. *Journal of Medical Internet Research, 21*(9), Article e14567. https://doi.org/10.2196/14567

Beierle, F., Probst, T., Allemand, M., Zimmermann, J., Pryss, R., Neff, P., Schlee, W., Stieger, S., & Budimir, S. (2020). Frequency and duration of daily smartphone usage in relation to personality traits. *Digital Psychology, 1*(1), 20–28. https://doi.org/10.24989/dp.v1i1.1821

Bell, L., Garnett, C., Qian, T., Perski, O., Williamson, E., & Potts, H. W. W. (2020). Engagement with a behavior change app for alcohol reduction: Data visualization for longitudinal observational study. *Journal of Medical Internet Research, 22*(12), Article e23369. https://doi.org/10.2196/23369

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Bucher, M. A., Suzuki, T., & Samuel, D. B. (2019). A meta-analytic review of personality traits and their associations with mental health treatment outcomes. *Clinical Pscychology Review, 70*, 51–63. https://doi.org/10.1016/j.cpr.2019.04.002

Canan, C. E., Flickinger, T. E., Waselewski, M., Tabackman, A., Baker, L., Eger, S., Waldman, A. L. D., Ingersoll, K., & Dillingham, R. (2021). Toward understanding the impact of mHealth features for people with HIV: A latent class analysis of PositiveLinks usage. *Translational Behavioral Medicine, 11*(1), 172–181. https://doi.org/10.1093/tbm/ibz180

De Vries, R. E., Ashton, M. C., & Lee, K. (2009). De zes belangrijkste persoonlijkheidsdimensies en de HEXACO Persoonlijkheidsvragenlijst. (The six most important personality dimensions and the HEXACO Personality questionnaire). *Gedrag en Organisatie, 22*, 232–274. https://doi.org/10.5117/2009.022.003.004

Donkin, L., & Glozier, N. (2012). Motivators and motivations to persist with online psychological interventions: A qualitative study of treatment completers. *Journal of Medical Internet Research, 14*(3), Article e91. https://doi.org/10.2196/jmir.2100

Dweck, C. S., & Yeager, D. S. (2019). Mindsets: A view from two eras. *Perspectives on Psychological Science, 14*, 81–96. https://doi.org/10.1177/1745691618804166

Galatzer-Levy, I. R., Brown, A. D., Henn-Haase, C., Metzler, T. J., Neylan, T. C., & Marmar, C. R. (2013). Positive and negative emotion prospectively predict trajectories of resilience and distress among high-exposure police officers. *Emotion, 13*(3), 545–553. https://doi.org/10.1037/a0031314

Garnett, C., Perski, O., Tombor, I., West, R., Michie, S., & Brown, J. (2018). Predictors of engagement, response to follow up, and extent of alcohol reduction in users of a smartphone app (drink less): Secondary analysis of a factorial randomized controlled trial. *JMIR mHealth and uHealth, 6*(12), Article e11175. https://doi.org/10.2196/11175

Hershfield, H. E., Garton, M. T., Ballard, K., Samanez-Larkin, G. R., & Knutson, B. (2009). Don't stop thinking about tomorrow: Individual differences in future self-continuity account for saving. *Judgm Decis Mak, 4*(4), 280–286.

Hollis, C., Falconer, C. J., Martin, J. L., Whittington, C., Stockton, S., Glazebrook, C., & Davies, E. B. (2017). Annual research review: Digital health interventions for children and young people with mental health problems – a systematic and meta-review. *Journal of Child Psychology and Psychiatry, 58*(4), 474–503. https://doi.org/10.1111/jcpp.12663

Kelders, S. M. (2015). *Involvement as a working mechanism for persuasive technology*. In T. MacTavish, & S. Basapur (Eds.). Springer. https://doi.org/10.1007/978-3-319-20306-5_1.

Kelders, S. M., Kip, H., & Greeff, J. (2020). Psychometric evaluation of the TWente Engagement with Ehealth Technologies Scale (TWEETS): Evaluation study. *J Med Internet Res, 22*, Article e17757. https://doi.org/10.2196/17757

Khwaja, M., Pieritz, S., Faisal, A. A., & Matic, A. (2021). Personality and engagement with digital mental health interventions. In *Proc ACM conference on user modeling, adaptation and personalization (UMAP '21)* (pp. 235–239). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3450613.3456823.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*(2), 329–358. https://doi.org/10.1207/s15327906mbr3902_8

Mertens, E. C. A., Van der Schalk, J., Siezenga, A. M., Gelder, Van, & J.-L. (2022a). Stimulating a futureoriented mindset and goal attainment through a smartphone-based intervention: Study protocol for a randomized controlled trial. *Internet Interv, 27*, 100509. https://doi.org/10.1016/j.invent.2022.100509

Mertens, E. C. A., Siezenga, A. M., Tettero, T., & Gelder, Van (2022b). A future orientation intervention delivered through a smartphone application and virtual reality: Study protocol for a randomized controlled trial. *BMC Psychology, 10*, 315. https://doi.org/10.1186/s40359-022-01025-x

Mertens, E. C. A., Siezenga, A. M., van der Schalk, J., & van Gelder, J.-L. (2024). A novel smartphonebased intervention aimed at increasing future orientation via the future self: A pilot randomized controlled trial of a prototype application. *Prevention science, 25*(2), 392–405. https://doi.org/10.1007/s11121-023-01609-y

Mikolasek, M., Witt, C. M., & Barth, J. (2018). Adherence to a mindfulness and relaxation self-care app for cancer patients: Mixed-methods feasibility study. *JMIR Mhealth Uhealth, 6*(12), Article e11271. https://doi.org/10.2196/11271

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*(6), 882–891. https://doi.org/10.1111/j.1530-0277.2000.tb02070.x

Nowak-Brzezińska, A., & Gaibei, I. (2022). How the outliers influence the quality of clustering? *Entropy, 24*(7), 917. https://doi.org/10.3390/e24070917

O'Brien, H. L. O., & Toms. (2009). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology, 59*(6). https://doi.org/10.1002/asi.20801

Perski, O., Blandford, A., West, R., & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine, 7*(2), 254–267. https://doi.org/10.1007/s13142-016-0453-1

Romano, K. A., Heron, K. E., Wilsey, C. N., Wright, S., & Cramer, R. J. (2022). Person-centered clustering of mHealth technology perceptions and usage motivations, and associations with physical and mental health correlates. *Journal of Technology in Behavioral Science, 7*, 167–178. https://doi.org/10.1007/s41347-021-00233-3

Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35–37). Windsor: UK: NFER-NELSON.

Siezenga, A. M., Mertens, E. C. A., & Van Gelder, J.-L. (2023). A look under the hood: Analyzing engagement and usage data of a smartphone-based intervention. *BMC Digit Health, 1*, 45. https://doi.org/10.1186/s44247-023-00048-7

Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., & Mani, M. (2015). Mobile app rating scale: A new tool for assessing the quality of health mobile apps. *JMIR mHealth uHealth, 3*(1), Article e27. doi:10.2196/mhealth.3422.

Su, J., Dugas, M., Guo, X., & Gao, G. (2020). Influence of personality on mHealth use in patients with diabetes: Prospective pilot study. *JMIR Mhealth Uhealth, 8*, Article e17709. https://doi.org/10.2196/17709

Van Gelder, J.-L., Luciano, E. C., Weulen Kranenbarg, M., & Hershfield, H. E. (2015). Friends with my future self: Longitudinal vividness intervention reduces delinquency. *Criminology, 53*(2), 158–179. https://doi.org/10.1111/1745-9125.12064

Van Ness, P. H., Murphy, T. E., Araujo, K. L., Pisani, M. A., & Allore, H. G. (2007). The use of missingness screens in clinical epidemiologic research has implications for regression modeling. *Journal of Clinical Epidemiology, 60*(12), 1239–1245. https://doi.org/10.1016/j.jclinepi.2007.03.006

Vannatta, R. A., & LaVenia, K. N. (2020). In P. Atkinson, S. Delamont, A. Cernat, J. W. Sakshaug, & R. A. Williams (Eds.), *Linear discriminant analysis*. https://doi.org/10.4135/9781526421036889610

Vinnikova, A., Lu, L., Wei, J., Fang, G., & Yan, J. (2020). The use of smartphone fitness applications: The role of self-efficacy and self-regulation. *International Journal of Environmental Research and Public Health, 17*(20), 7639. https://doi.org/10.3390/ijerph17207639

Yardley, L., Morrison, L., Bradbury, K., & Muller, I. (2015). The person-based approach to intervention development: Application to digital health-related behavior change interventions. *Journal of Medical Internet Research, 17*(1), e30. https://doi.org/10.2196/jmir.4055