A look under the hood: analyzing engagement and usage data of a smartphone-based intervention

Aniek M. Siezenga[1,2]📷, Esther C. A. Mertens[1]📷 and Jean-Louis van Gelder[1,2]*📷

Abstract

Background Engagement with smartphone-based interventions stimulates adherence and improves the likelihood of gaining benefits from intervention content. Research often relies on system usage data to capture engagement. However, to what extent usage data reflect engagement is still an open empirical question. We studied how usage data relate to engagement, and how both relate to intervention outcomes.

Methods We drew data from a randomized controlled trial (RCT) ($N=86$) evaluating a smartphone-based intervention that aims to stimulate future self-identification (i.e., future self vividness, valence, relatedness). General app engagement and feature-specific engagement were retrospectively measured. Usage data (i.e., duration, number of logins, number of days used, exposure to intervention content) were unobtrusively registered.

Results Engagement and usage data were not correlated. Multiple linear regression analyses revealed that general app engagement predicted future self vividness ($p=.042$) and relatedness ($p=.004$). Furthermore, engagement with several specific features also predicted aspects of future self-identification ($p=.005 – .032$). For usage data, the number of logins predicted future self vividness ($p=.042$) and exposure to intervention content predicted future self valence ($p=.002$).

Conclusions Usage data did not reflect engagement and the latter was the better predictor of intervention outcomes. Thus, the relation between usage data and engagement is likely to be intervention-specific and the unqualified use of the former as an indicator of the latter may result in measurement error. We provide recommendations on how to capture engagement and app use in more valid ways.

Keywords Smartphone app, mHealth intervention, Engagement, Usage data, Log-data, Future Self-Identification

Smartphone-based interventions are easy to access, scalable, user-driven, always available, and have shown positive effects in increasing mental health and wellbeing [1]. However, users must be "engaged" with the smartphone application (app) in order to reap its benefits [2]. Engagement has been defined as the affective (i.e., feelings associated with making progress), cognitive (e.g., feeling supported or motivated to reach goals) and behavioral (i.e., embedded app usage in daily routines) user experience

*Correspondence:
Jean-Louis van Gelder
j.vangelder@csl.mpg.de
[1] Department of Education and Child Studies, Leiden University, Wassenaarseweg 52, 2333AK Leiden, The Netherlands
[2] Department of Criminology, Max Planck Institute for the Study of Crime, Security and Law, Günterstalstraße 73, 79100 Freiburg im Breisgau, Germany

[3]. App engagement is typically assessed with system usage data [4], under the assumption that app usage reflects engagement [5, 6]. However, empirical evidence for this assumption is both limited and mixed [7]. In the present study, we examined to what extent usage data can be used as a proxy for engagement in an app intervention that aims to stimulate future self-identification. We also examined to what extent engagement and usage data are related to intervention outcomes. Increasing our understanding of how engagement and app use contribute to outcomes can aid the design and development of app interventions, and, potentially, result in more effective interventions.

Usage data: A proxy for engagement?

Even though studies often employ usage data to reflect engagement (e.g., [4]), the relation between both may be less straightforward than commonly assumed. First, it can be difficult to infer meaning from usage data as the processes behind the app use are unknown. This is known as the "black box phenomenon" [8]. To illustrate, app use can be effective in the sense that it contributes to achieving the aims of the app. In this article, we will refer to this as "effective use". However, people can also use the technology ineffectively by using the app inattentively, such as using the app whilst being distracted by friends or whilst simultaneously being engaged in another activity (henceforth: "ineffective use"). Use that is registered by the technology even though the user is not exposed to the intervention content can also occur (henceforth: "non-use"). Non-use can, for example, occur when users fail to close the app, or when someone else other than the intended user is using the app. The black box phenomenon makes it difficult to infer type of use, i.e., effective use, ineffective use, or non-use from usage data.

Second, even when usage data reflect effective use, the extent to which these data reflect engagement is questionable. After all, users who do not feel engaged with the app could nevertheless be effectively using it out of necessity. For example, a fitness app prescribed by one's doctor to improve a healthy lifestyle may be used frequently (thus, resulting in high usage) out of a sense of duty or perceived necessity, but might not be experienced as engaging. Thus, treating usage data as a proxy for engagement neglects the fact that app use may not necessarily mean effective use, and neglects individual reasons for the app use. Therefore, the relationship between engagement and usage data requires further investigation.

To our knowledge, only two studies examined how usage data relate to engagement. Yeager and Benight [9] found that one out of four usage data variables in their study loaded on a factor representing engagement, and concluded from this finding that usage data can be seen as measure of engagement. However, the extent to which usage data captured engagement in this study is questionable, as three out of four usage variables did not load on the latent engagement construct and the one variable that did, had a low factor loading. Perski et al. [10] evaluated the psychometric properties of their Digital Behavior Change Interventions Engagement Scale designed to assess engagement with digital behavior change interventions. The total scale score correlated weakly, but significantly, with only one out of three usage variables, and did not significantly predict the number of subsequent logins. Therefore, strong evidence regarding a relationship between engagement and usage data is lacking.

Engagement, usage data, and intervention outcomes

Prior work concluded that feeling engaged with an app intervention is related to the intervention's outcomes [3, 11]. However, this conclusion is largely based on studies relying on usage data as proxies for engagement [4, 12]. Only few studies included subjective measures of engagement to examine this relationship [13]. Therefore, the importance of engagement for gaining benefit from the intervention is less clear than often assumed and needs to be studied further.

Regarding the relationship between usage data and intervention outcomes, in general, usage data appear to be positively related to outcomes [6]. However, not all studies support this positive relation (e.g., [14]). These mixed findings might be the result of research often using general types of usage data (e.g., total time spent in the app) rather than specific ones (i.e., usage data that capture feature-specific app usage) or composite usage variables (e.g., number of user actions per login) [6]. General usage variables are more susceptible to the black box phenomenon, as they contain less meaning and can differentiate less well between effective use, ineffective use, and non-use. For example, a general usage measure, as total app use duration, would fail to differentiate between one user accessing three features in 20 min over the course of multiple days, and another user accessing six features in 20 min on one day. In order to better understand how different types of app use relate to intervention outcomes, research needs to include both specific usage variables and composite usage variables.

Present study

In this study, we examined 1) the association between engagement and usage data, and 2) how both predict intervention outcomes. We drew data from a randomized controlled trial (RCT) evaluating a smartphone-based intervention, FutureU, which aims to stimulate future-oriented thinking and personal goal attainment through strengthening future self-identification (i.e., feelings of

Siezenga *et al. BMC Digital Health* (2023) 1:45

Page 3 of 10

vividness, valence, and relatedness towards the future self) (see [15]). We assessed engagement on two levels: 1) engagement with the app in general, i.e., general app engagement, and 2) engagement with specific features of the app. Regarding usage data, we analyzed two general types of data that are often used in other studies and are more likely to be influenced by the black box phenomenon (i.e., total time spent in the app, and the percentage of days the app was used), and two types of usage data that reflect use of features deemed relevant for the app's aim (i.e., use of the intervention technique at each login, and use of app content). Regarding the first aim, taking findings from prior research into consideration, we expected a significant positive correlation between all types of engagement and usage data we assessed. Regarding the second aim, we expected that more engagement with, and use of, the app increases future self-identification significantly.

Method

Study design, participants, and procedure

We employed data from the intervention condition of a two-armed RCT (October 2021 – March 2022), given that participants in the control condition did not use an app. Ethical approval was obtained by the Leiden University's Ethics Board. The RCT was prospectively registered in the Netherlands Trial Register, number: NL9671 (see [15] for the study protocol).

Participants were first-year university students ($N=87$). One participant was excluded from analyses due to an installation error. Participants ($n_{female}=70$; 81.40%) were on average 19.72 years old ($SD=3.44$) and studied social sciences ($n=81$; 94.20%).

Participants started with an intake session at the university during which they gave written informed consent, completed the pre-measurement questionnaire (T1),[1] took a picture of their face (i.e., a "selfie") for avatar creation, and installed the FutureU app. Participants were instructed to interact with the app on a daily basis. Immediately following the 21-day intervention period, participants completed the post-measurement questionnaire (T4). When participants did not respond to the post-measurement questionnaire, they were considered drop-outs. When participants did not complete the post-measurement questionnaire in time (i.e., within 8 days), their responses were considered missing data.

---

[1] During the intervention, the Netherlands went into COVID-lockdown (December 19 2021 to January 26 2022). MANOVA ($F(3,83)=2.26$, $p=.088$; partial $\eta^2=.075$) revealed no significant differences between participants starting before and during lockdown in demographic variables and T1 and T4 vividness, valence and relatedness.

Participants received compensation (6 study credits or 25 euro) for completing the questionnaires.

FutureU app

FutureU introduces users to an avatar that represents their future self. This is a three-dimensional age-progressed avatar, based on the participant's selfie. The app is built around a scripted chat conversation in which users interact with their future self (see Fig. 1B). The future self provides psychoeducation, asks questions, sends motivational messages and explains the assignments. Users can reply to the messages and answer questions in the chat. Besides this chat-feature, the FutureU app contains other features, such as a feature to take the perspective of the future self on a current challenge (referred to as the time travel portal) and a feature focused on goal attainment (referred to as the goal attainment method). To login onto the app, users need to touch the future self's virtual finger to unblur the screen and make the future self visible (see Fig. 1A). This is referred to as the connection mechanic, and functions as an intervention technique to increase the vividness of the future self by exposure to the future self [16]. When users skip a day or abort a session, they will receive the missed content the next time they interact with the app (see the study protocol [15] for more information about the intervention's content).

The intervention consists of three week-long modules, each containing different intervention components. The first module aims to promote future self-identification. The second module includes perspective taking exercises for which participants use the built-in time travel portal. The third module aims to stimulate a growth mindset [17] and goal achievement by teaching users Mental Contrasting and Implementation Intentions (MCII; [18]). Different module features (e.g., the time travel portal) can be accessed via the home menu (see Fig. 1C). Daily interactions last approximately 2 to 5 min. For a more extensive overview of module content, see [15].

Measurements

Engagement

All engagement items were measured at T4 on 7-point Likert scales, ranging from 1=Completely disagree to 7=Completely agree (see Additional file 1). General app engagement was measured with a shortened version of the Mobile App Rating Scale Section F (MARS; $n=3$, $\alpha=0.88$, e.g., "This app increases my knowledge of my future self", [19]). Engagement with the app's specific features was measured using a self-designed survey. Specifically, we assessed participants' opinion of, and engagement with their future self avatar ($n=4$, $\alpha=0.85$, e.g., "I

Fig. 1 Impressions of the FutureU App showing the (**A**) Connection Mechanic, **B** Chat, and (**C**) Home Menu

recognized myself in the avatar"), the time travel portal ($n = 5$, $\alpha = 0.88$, e.g., "I was able to take the perspective of my future self in the time travel portal"), and the goal attainment module ($n = 3$, $\alpha = 0.95$, e.g., "I thought it was useful to practice the goal attainment method").

App usage

Usage data were registered by the app and stored on a protected web server. These data contained the date,

and start and end time of each login. From these data, we computed two general types of usage variables: 1) the total number of minutes that users spent using the app ("Duration") and 2) the percentage of days that users interacted with the app during the 21-day intervention period ("Days").

Furthermore, based on the usage data we computed two variables that reflected use of specific app features: 1) the number of times that users accessed the app (i.e., logged-in) ("Logins") and 2) percentage of content

Siezenga *et al. BMC Digital Health*   (2023) 1:45

Page 5 of 10

exposure[2] ("Exposure to intervention content"). We computed exposure to intervention content taking the following four steps. First, for each day, we timed the minimum duration for delivering the scripted messages in the chat. Second, we subtracted this minimum duration from the time that participants spent in the app that day. Negative values indicate that users spent less time in the app than the time it would take to deliver all scripted chat messages, meaning that they were not exposed to all intended intervention content of the concerned day. Positive values meant that users spent more time in the app than the minimum duration of the chat, meaning that they were exposed to the complete content of that day. Third, we categorized each value, representing 0 = No exposure (i.e., no login that day), 1 = Spending one to five seconds in the app (i.e., mere interaction with the connection mechanic), 2 = Exposure to some of the day's intervention content, but less than half of the content, 3 = Exposure to more than half of the day's content, but less than the full content, 4 = Full exposure. To illustrate, the minimum chat duration of day 3 was 178 s. For a user spending 50 s in the app, day 3 was recoded into category 2 (less than half exposure). Last, we computed the sum of these category values for the 21-day intervention period (maximum sum score is 84, i.e., 21 days a score of 4), which we transformed into a percentage representing the participant's exposure to the complete intervention content (e.g., a sum score of 84 represents 100% exposure to the intervention content).

Intervention outcomes

Future Self-Identification (FSI) was assessed at baseline and post-measurement with three scales.

Vividness of the future self indicated the degree to which people have a vivid image of their future self, which was assessed with five items (α = 0.95, e.g., "I find it hard to imagine what kind of person I will be in 10 years from now" [20]) 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 the future" [21]) using a self-assessment manikin, ranging from 1 = negative to 9 = positive.

Relatedness to the future self, measured with two items (α = 0.79) [21], indicated how connected and similar participants feel to their future self. Participants rated the extent to which two circles, representing the present

Table 1 Descriptive statistics of engagement, usage data, and intervention outcomes

Variable	*n*	*M*	*SD*	Min	Max
Engagement					
MARS	82	4.27	1.52	1.00	7.00
TTP	82	3.48	1.35	1.00	6.20
GAM	82	4.56	1.38	1.00	7.00
Avatar	82	3.79	1.44	1.00	6.25
Usage data					
Duration	86	62.91	28.02	5.82	210.70
Logins	86	28.63	10.82	2	57
Days	86	69.16	22.55	9.52	100
Exposure	86	83.44	18.78	20.24	100
Outcomes T4					
Vividness	82	3.83	1.45	1.00	7.00
Valence	82	6.45	1.35	3.00	7.00
Relatedness	82	4.16	1.07	1.50	7.00

MARS Mobile App Rating Scale, *TTP* Time Travel Portal, *GAM* Goal Achievement Method

and future self, overlap on a 7-point scale ranging from 1 = no overlap to 7 = almost complete overlap.

Results

All analyses were performed using IBM SPSS Statistics v28 (IBM Corp). We considered $p < 0.05$ to be statistically significant.

Attrition and descriptive statistics

At post-measurement, one participant was a drop-out due to a non-response on the post-measurement questionnaire. One participant failed to fill in the post-measurement survey on time and therefore this survey response was treated as missing data. Two participants partly completed the questionnaire, which also resulted in missing data. Missing data were handled with listwise deletion, since the percentage of missing data was low (5.7%) and Little's MCARs test showed a random pattern of missing data ($\chi(4)^2 = 3.92$, $p = 0.417$).

See Table 1 for descriptive statistics of subjective engagement, usage data, and intervention outcomes. Over the course of the intervention period a decline in app use was visible, as the average number of missed days increased (Module 1 M = 0.87 missed days, $n = 86$; Module 2 M = 2.36 missed days, $n = 82$; Module 3 M = 3.24 missed days, $n = 78$[3]). The median rate of exposure to

---

[2] Since missed intervention content would be shown during a next login, the number of days would not be a valid measure of exposure to the intervention's content. See Additional file 2 for an extended explanation of the exposure variable.

[3] There were no significant differences between the eight participants that did not login during Module 2 and/or Module 3 and the 78 participants that did, in terms of age, gender, education, and baseline vividness, relatedness, and valence (see Additional file 3).

intervention content was 90.48%. Twelve participants had been exposed to the entire intervention content after the 21-day intervention period.

Correlation engagement and usage data

We assessed the correlation between app engagement and usage data with Spearman correlation analyses. The results showed that all four engagement subscales correlated significantly among each other, as did all usage variables (see Table 2). No significant correlations emerged between engagement and usage variables.

Regression analyses engagement, usage data, and future self-identification

We assessed the relationship between app engagement and intervention outcomes with multivariate linear regressions in which we controlled for baseline vividness, valence and relatedness (see Table 3). The results showed that general app engagement was positively related to vividness of and relatedness to the future self, with respectively a small to medium ($\beta = 0.16$) and a strong effect size ($\beta = 0.26$). Engagement with the time travel portal was positively related to an increase in feeling related to the future self with a medium effect size ($\beta = 0.19$).

Table 2 Descriptive statistics and Spearman correlations for engagement ($n = 82$) and usage data ($n = 86$)

Variable	1	2	3	4	5	6	7	8
1. MARS	—							
2. TTP	.62**	—						
3. GAM	.43**	.41**	—					
4. Avatar	.49**	.33*	.43**	—				
5. Duration	.11	.03	.04	.04	—			
6. Logins	.01	-.05	-.09	-.11	.53**	—		
7. Days	.04	-.08	-.11	-.00	.62**	.67**	—	
8. Exposure	-.04	-.17	.00	-.02	.74**	.48**	.79**	—

MARS Mobile App Rating Scale, *TTP* Time Travel Portal, *GAM* Goal Achievement Method

*$p < .01$, **$p < .001$

Table 3 Multivariate regression analyses: MARS, TTP, GAM, avatar predicting FSI

Effect	*b*	*SE*	β	*t*	95% CI		*p*
					LL	*UL*	
MARS							
Vividness	.16	.08	.16	2.07	0.01	0.31	.042
Valence	.09	.09	.10	1.02	-0.08	0.26	.310
Relatedness	.18	.06	.26	3.00	0.06	0.31	.004
TTP							
Vividness	.12	.09	.11	1.38	-0.05	0.30	.171
Valence	.03	.10	.03	0.28	-0.17	0.22	.779
Relatedness	.15	.07	.19	2.18	0.01	0.30	.032
GAM							
Vividness	.08	.09	.07	0.88	-0.10	0.25	.383
Valence	.09	.09	.09	0.97	-0.10	0.28	.334
Relatedness	.13	.07	.17	1.89	-0.01	0.27	.063
Avatar							
Vividness	.18	.08	.18	2.24	0.20	0.34	.028
Valence	.25	.09	.27	2.89	0.08	0.42	.005
Relatedness	.12	.07	.16	1.76	-0.02	0.25	.082

n = 82, *FSI* Future self-identification, *MARS* Mobile App Rating Scale, *TTP* Time Travel Portal, *GAM* Goal Achievement Method, *CI* confidence interval, *LL* lower limit, *UL* upper limit

Siezenga *et al. BMC Digital Health*   (2023) 1:45

Page 7 of 10

Table 4 Multivariate regression analyses: duration, logins, days, exposure predicting FSI

Effect	*b*	*SE*	β	*t*	95% CI		*p*
					LL	UL	
Duration							
Vividness	.01	.00	.11	1.31	-0.00	0.01	.195
Valence	.01	.01	.08	0.80	-0.01	0.01	.301
Relatedness	.00	.00	.09	0.93	-0.00	0.01	.357
Logins							
Vividness	.02	.01	.16	2.06	0.00	0.04	.042
Valence	.01	.01	.11	1.19	-0.01	0.04	.237
Relatedness	.01	.01	.11	1.21	-0.01	0.03	.231
Days							
Vividness	.01	.01	.13	1.64	-0.00	0.02	.105
Valence	.01	.01	.15	1.62	-0.00	0.02	.108
Relatedness	.01	.00	.16	1.72	-0.00	0.02	.090
Exposure							
Vividness	.01	.01	.07	0.89	-0.01	0.02	.379
Valence	.02	.01	.29	3.27	0.01	0.03	.002
Relatedness	.01	.01	.16	1.67	-0.00	0.02	.099

n = 86, *FSI* future self-identification, *CI* confidence interval, *LL* lower limit, *UL* upper limit

Engagement with the future self's avatar was positively related to an increase in vividness of and valence towards the future self, with respectively a medium (β = 0.18) and a strong (β = 0.27) effect size.

Second, we assessed the relationship between usage data and intervention outcomes. The multivariate linear regressions (see Table 4) revealed that number of logins positively predicted an increase in vividness of the future self with a small to medium effect size (β = 0.16). Furthermore, exposure to intervention content positively predicted valence of the future self with a strong effect size (β = 0.29). Duration and days did not predict any of the dependent variables.[4]

Discussion

We examined how engagement and usage data relate to each other, and how each separately relates to intervention outcomes of a smartphone-based intervention. First, our results showed that engagement and usage data were unrelated as no significant correlations between engagement and usage data emerged. Second, higher levels of engagement were related to stronger intervention outcomes. More specifically, general app engagement predicted increases in vividness and relatedness, engagement with the time travel portal predicted an increase in relatedness, and engagement with the future self-avatar predicted increases in vividness and valence. Regarding app use, we found that specific usage data (i.e., logins) and composite usage variables (i.e., exposure to intervention content) were related to intervention outcomes, but general usage variables (i.e., time spent in the app and percentage of days the app was used) were not. Specifically, the number of logins predicted an increase in vividness, and exposure to intervention content predicted valence.

Although app engagement and app use might indeed be unrelated [3], a lack of association between the two may also be an indication that most usage data are inadequate measures of engagement. As usage data generally do not reveal the process behind the use of the app (i.e., the black box phenomenon), it is difficult to differentiate between effective use, ineffective use, and non-use. As a consequence, it is not surprising that usage data were unrelated to levels of engagement. Hypothetically, participants could open the app and put their phone down without closing the app, resulting in usage data representing high usage without them actually interacting with the app. Based on this scenario, a relation between usage data and engagement would not be expected. Therefore, further investigation in how to unravel the processes

---

[4] To assess the robustness of our findings, we ran sensitivity analyses. All main analyses were rerun with two subsamples of participants. In the first subsample, 10 participants who experienced technical problems with the app (e.g., the future self-avatar could not be created) were excluded. In the second subsample, 14 participants were excluded who failed one or two attention checks that were present in the post-measurement questionnaire, instructing participants to select a specific answer [22]. The sensitivity analyses showed similar patterns as the main analyses (see Additional file 4) based on the parameters and associations between the dependent and independent variables (as *p*-values are affected by the reduced sample sizes).

behind usage data is needed before firm conclusions can be drawn about the associations between engagement and usage data.

While it remains a challenge to link general usage data to engagement due to the black box phenomenon, usage data do seem useful to analyze within an intervention context when they capture use of specific intervention components. For example, our findings showed that number of logins predicted intervention outcomes. Even though number of logins seems like a general type of usage data, in the FutureU app each login meant using the connection mechanic (i.e., one of the intervention techniques). Therefore, this variable actually captured specific usage of an intervention component. Similarly, exposure to the intervention content, and not total duration, predicted intervention outcomes. As the exposure variable captured usage more validly compared to total duration, the black box effect was reduced. Not only do both findings indicate the potential of the FutureU intervention app, it also emphasizes the potential of usage data when considered carefully: More use of the theory-based intervention components predicted stronger intervention outcomes. Hence, specific usage data seem to be able to capture participants' use of intervention components, enabling the investigation of which intervention components are effective. A similar finding also emerged in Donkin et al. [23], who found that number of completed activities (i.e., intervention components) per login of a theory-based internet intervention was significantly associated with stronger intervention outcomes, where general usage data were not.

Regarding the relationship between engagement and intervention outcomes, we found a generally consistent pattern that both general app engagement and engagement with app-specific features predicted outcomes. These results are in line with previous research finding a positive relationship between subjectively reported engagement and intervention outcomes (e.g., [3, 11]) and stress the importance of feeling engaged in a smartphone-based intervention in order to optimize intervention effects.

Limitations

Our results must be viewed in light of the following limitations. First, as participants were randomly appointed to the experimental condition, they might have been less intrinsically motivated to use the app compared to users who would download the app out of interest or need for support. As intrinsic motivation to use an app intervention is important for both engagement and app use [24], the results might be different if we would have released the app "in the wild". Nevertheless, the FutureU app showed relatively high retention and exposure rates compared to other smartphone-based interventions (e.g., on average 3.9% of participants open an app intervention on day 15 [25] versus 60.5% of participants opening the FutureU app on day 15). This could have been the result of the foreseeable FutureU intervention period of three weeks, and the short and clearly defined daily interactions. Additionally, although participants received a compensation for completing questionnaires and not for using the app, this compensation could have indirectly affected our retention rate. Second, we ran multiple regression analyses to avoid multicollinearity within the analyses. This could have inflated the alpha, but as we were interested in discovering patterns instead of specific results, we decided not to correct for multiple comparisons. Third, engagement was measured at post-measurement. To reduce recall bias and gain more insight in the expected natural fluctuations in engagement [26], we recommend to regularly assess engagement during intervention periods. Fourth, to assess app engagement, we used a self-developed survey, as most studies assessing engagement [27]. Therefore, comparability of our findings with other studies is limited. In addition to measures that capture feature-specific engagement, standardized questionnaires for assessing general app engagement could be included in research to be able to generalize findings (e.g., [3]).

Lastly, we find it important to note that our findings may be app and/or intervention specific. That is, since app interventions are diverse in terms of their aims, content, complexity, design, etc. [4], it is challenging to compare results between studies and we stress the need for replication of our findings among different app interventions and target groups.

Conclusion

The current study is the first to explicitly study the association between engagement with and use of a smartphone-based intervention. Even though a positive relationship between the two is widely assumed, our results revealed a lack of relationships between engagement and usage data. Thus, engagement cannot simply be inferred from usage data. As engagement was important in predicting intervention outcomes, we recommend measuring it using survey instruments instead of usage data. Although usage data cannot straightforwardly be used as proxy for engagement, they can nevertheless provide valuable insights. For example, usage data can be employed to examine how intervention components have been used by participants and how this use relates to intervention outcomes. To this end, we emphasize the importance of considering more sophisticated usage variables (e.g., through constructing more detailed composite variables) and to include

Siezenga *et al. BMC Digital Health*　　(2023) 1:45

Page 9 of 10

specific usage data (e.g., time spent interacting with a specific feature), rather than relying on general usage data. It is worth noting, however, that even with the most specific types of usage data, it remains a challenge to differentiate between effective use, ineffective use, and non-use, due to the black box phenomenon. Future studies could aim to differentiate between the three types of use for which the potential of advanced analysis methods, such as natural language processing or artificial intelligence, could be explored. Furthermore, the relationship between engagement, usage data, and intervention outcomes requires replication in different smartphone-based interventions. Until then, future research should be careful to include usage data as proxy for engagement.

Abbreviations

RCT	Randomized Controlled Trial
App	Smartphone application
MCII	Mental Contrasting and Implementation Intentions
MANOVA	Multivariate analysis of variance
MARS	Mobile App Rating Scale
FSI	Future Self-Identification
TTP	Time Travel Portal
GAM	Goal Achievement Method
CI	Confidence interval
LL	Lower Limit
UL	Upper Limit

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s44247-023-00048-7.

> **Additional file 1: Table S1.1.** Engagement Survey. **Table S1.2.** Model Fit Statistics. **Figure S1.1.** Four-Factor Model of the Engagement Survey.
>
> **Additional file 2.** Computation of the Exposure Variable.
>
> **Additional file 3.** Attrition of App Users.
>
> **Additional file 4: Table S4.1.** Sensitivity Analysis 1 of Research Question 1. **Table S4.2.** Sensitivity Analysis 2 of Research Question 1. **Table S4.3.** Sensitivity Analysis 1 of Research Question 2. **Table S4.4.** Sensitivity Analysis 2 of Research Question 2.

Availability of data and materials

The dataset analyzed during the current study are available after publication in the Center for Open Science (https://osf.io/gct8e/?view_only=ddb199078d7641fa8a4f648bfc9a8395).

Declarations

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 (ECPW-2021/320). 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.

Competing interests

The authors declare no competing interests.

References

1. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: A meta-analysis of randomized controlled trials. World Psychiatry. 2019;18:325–36. https://doi.org/10.1002/wps.20673.
2. Kelders SM, Van Zyl LE, Ludden GDS. The concept and components of engagement in different domains applied to eHealth: A systematic scoping review. Front Psychol. 2020;11:926. https://doi.org/10.3389/fpsyg.2020.00926.
3. Kelders SM, Kip H, Greeff J. Psychometric evaluation of the TWente Engagement with Ehealth Technologies Scale (TWEETS): Evaluation study. J Medical Internet Res. 2020;22:17757. https://doi.org/10.2196/17757.
4. Bijkerk LE, Oenema A, Geschwind N, Spigt M. Measuring engagement with mental health and behavior change interventions: An integrative review of methods and instruments. Int J Behav Med. 2023;30:155–66. https://doi.org/10.1007/s12529-022-10086-6.
5. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from critical interpretive synthesis. Transl Behav Med. 2017;7:254–67. https://doi.org/10.1007/s13142-016-0453-1.
6. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: Scoping review. JMIR mHealth uHealth. 2019;7:11941. https://doi.org/10.2196/11941.
7. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med. 2016;51:833–42. https://doi.org/10.1016/j.amepre.2016.06.015.
8. Sieverink F, Kelders S, Poel M, Van Gemert-Pijnen L. Opening the Black Box of electronic health: Collecting, analyzing and interpreting log data. JMIR Res Protoc. 2017;6:156. https://doi.org/10.2196/resprot.6452.
9. Yeager CM, Benight CC. Engagement, predictors, and outcomes of a trauma recovery digital mental health intervention: longitudinal study. JMIR Ment Health. 2022;9:35048. https://doi.org/10.2196/35048.
10. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): Development and psychometric evaluation of the "DBCI engagement scale." Transl Behav Med. 2020;10:267–77. https://doi.org/10.1093/tbm/ibz039.
11. Graham AK, Kwasny MJ, Lattie EG, Greene CJ, Gupta NV, Reddy M, et al. Targeting subjective engagement in experimental therapeutics for digital mental health interventions. Internet Interv. 2021;25:100403. https://doi.org/10.1016/j.invent.2021.100403.
12. Short CE, Desmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change

interventions: Viewpoint of methodologies. J Medical Internet Res. 2018;20:292. https://doi.org/10.2196/jmir.9397.

13. Oakley-Girvan I, Yunis R, Longmire M, Schwartz OJ. What works best to engage participants in mobile app interventions and e-Health: A scoping review. Telemed E-Health. 2022;28:768–80. https://doi.org/10.1089/tmj.2021.0176.

14. Henry JA, Thielman E, Zaugg T, Kaelin C, Choma C, Chang B, et al. Development and field testing of a smartphone "app" for tinnitus management. Int J Audiol. 2017;56:784–92. https://doi.org/10.1080/14992027.2017.1338762.

15. Mertens ECA, Van der Schalk J, Siezenga AM, Van Gelder J-L. Stimulating a future-oriented mindset and goal attainment through a smartphone-based intervention: Study protocol for a randomized controlled trial. Internet Interv. 2022;27:100509. https://doi.org/10.1016/j.invent.2022.100509.

16. McMichael SL, Bixter MT, Okun MA, Bunker CJ, Graudejus O, Grimm KJ, et al. Is seeing believing? A longitudinal study of vividness of the future and its effects on academic self-efficacy and success in college. Pers Soc Psychol Bull. 2022;48:478–92. https://doi.org/10.1177/01461672211015888.

17. Dweck CS, Yeager DS. Mindsets: A view from two eras. Perspect Psychol Sci. 2019;14:81–96. https://doi.org/10.1177/1745691618804166.

18. Oettingen G, Gollwitzer PM. Strategies of setting and implementing goals. Mental contrasting and implementation intentions. In: Maddux JE, Tangney JP, editors. Social psychological foundations of clinical psychology. New York: The Guilford Press; 2010; 2010. p. 114–35.

19. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: A new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth. 2015;3:27 (http://mhealth.jmir.org/2015/1/e27/).

20. Van Gelder J-L, Luciano EC, Weulen Kranenbarg M, Hershfield HE. Friends with my future self: Longitudinal vividness intervention reduces delinquency. Criminology. 2015;53:158–79. https://doi.org/10.1111/1745-9125.12064.

21. Hershfield HE, Garton MT, Ballard K, Samanez-Larkin GR, Knutson B. Don't stop thinking about tomorrow: Individual differences in future self-continuity account for saving. Judgm Decis Mak. 2009;4:280–6.

22. Gummer T, Roßmann J, Silber H. Using instructed response items as attention checks in web surveys: Properties and implementation. Sociol Methods Res. 2018;50:238–64. https://doi.org/10.1177/0049124118769083.

23. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Medical Internet Res. 2011;13:52. https://doi.org/10.2196/jmir.1772.

24. Doherty K, Doherty G. Engagement in HCI: Conception, theory and measurement. ACM Comput Surv. 2018;51:1–39. https://doi.org/10.1145/3234149.

25. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: Systematic search and panel-based usage analysis. J Medical Internet Res. 2019;21:14567. https://doi.org/10.2196/14567.

26. O'Brien HLO, Morton E, Kampen A, Barnes SJ, Michalak EE. Beyond clicks and downloads: A call for a more comprehensive approach to measuring mobile-health app engagement. Br J Psychiatry. 2020;6:86. https://doi.org/10.1192/bjo.2020.72.

27. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: A review of measurement, reporting and validity. Psychiatr Serv. 2019;70:538–44. https://doi.org/10.1176/appi.ps.201800519.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.