



Perski, O., Lumsden, J., Garnett, C., Blandford, A., West, R., & Michie, S. (2019). Assessing the psychometric properties of the digital behavior change intervention engagement scale in users of an app for reducing alcohol consumption: Evaluation study. *Journal of Medical Internet Research*, 21(11), Article e16197. <https://doi.org/10.2196/16197>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.2196/16197](https://doi.org/10.2196/16197)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via JMIR at <https://www.jmir.org/2019/11/e16197/>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Original Paper

Assessing the Psychometric Properties of the Digital Behavior Change Intervention Engagement Scale in Users of an App for Reducing Alcohol Consumption: Evaluation Study

Olga Perski¹, PhD; Jim Lumsden², PhD; Claire Garnett¹, PhD; Ann Blandford³, PhD; Robert West¹, PhD; Susan Michie⁴, DPhil

¹Department of Behavioural Science and Health, University College London, London, United Kingdom

²UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, United Kingdom

³UCL Interaction Centre, University College London, London, United Kingdom

⁴Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom

Corresponding Author:

Olga Perski, PhD

Department of Behavioural Science and Health

University College London

1-19 Torrington Place

London, WC1E 7HB

United Kingdom

Phone: 44 20 7679 1258

Email: olga.perski@ucl.ac.uk

Abstract

Background: The level and type of engagement with digital behavior change interventions (DBCIs) are likely to influence their effectiveness, but validated self-report measures of engagement are lacking. The DBCI Engagement Scale was designed to assess behavioral (ie, amount, depth of use) and experiential (ie, attention, interest, enjoyment) dimensions of engagement.

Objective: We aimed to assess the psychometric properties of the DBCI Engagement Scale in users of a smartphone app for reducing alcohol consumption.

Methods: Participants (N=147) were UK-based, adult, excessive drinkers recruited via an online research platform. Participants downloaded the *Drink Less* app and completed the scale immediately after their first login in exchange for a financial reward. Criterion variables included the objectively recorded amount of use, depth of use, and subsequent login. Five types of validity (ie, construct, criterion, predictive, incremental, divergent) were examined in exploratory factor, correlational, and regression analyses. The Cronbach alpha was calculated to assess the scale's internal reliability. Covariates included motivation to reduce alcohol consumption.

Results: Responses on the DBCI Engagement Scale could be characterized in terms of two largely independent subscales related to experience and behavior. The experiential and behavioral subscales showed high ($\alpha=.78$) and moderate ($\alpha=.45$) internal reliability, respectively. Total scale scores predicted future behavioral engagement (ie, subsequent login) with and without adjusting for users' motivation to reduce alcohol consumption (adjusted odds ratio [OR_{adj}]=1.14; 95% CI 1.03-1.27; $P=.01$), which was driven by the experiential (OR_{adj}=1.19; 95% CI 1.05-1.34; $P=.006$) but not the behavioral subscale.

Conclusions: The DBCI Engagement Scale assesses behavioral and experiential aspects of engagement. The behavioral subscale may not be a valid indicator of behavioral engagement. The experiential subscale can predict subsequent behavioral engagement with an app for reducing alcohol consumption. Further refinements and validation of the scale in larger samples and across different DBCIs are needed.

(*J Med Internet Res* 2019;21(11):e16197) doi: [10.2196/16197](https://doi.org/10.2196/16197)

KEYWORDS

engagement; digital behaviour change interventions; mHealth; psychometrics; self-report scale; smartphone apps; excessive alcohol consumption

Introduction

Some level of engagement with digital behavior change interventions (DBCIs) is necessary for the effectiveness of such interventions [1]. However, observed levels of engagement with DBCIs are often considered too limited to support behavior change [2]. For example, a systematic review of Web-based health interventions found that approximately 50% of participants engaged with the interventions in the desired manner, with estimates varying between 10% and 90% across trials [3]. Studies conducted across different settings and target behaviors report a positive association of DBCI engagement and intervention effectiveness [4,5], suggesting that these variables may be linked via a dose-response function [1,6]. However, it is also plausible that individuals who are more successful in achieving change in the behavior targeted by the DBCI engage with DBCIs more [7] or that a limited amount of engagement is sufficient for bringing about meaningful change in some users (ie, “effective engagement”) [6]. Attempts have been made to characterize the function linking engagement with intervention effectiveness [1,7-9], but progress is hindered due to the use of different definitions and measures of engagement across studies.

The question of what it means for someone to be engaged with a DBCI has been of interest to psychologists and computer scientists alike. Broadly, psychologists have defined engagement as the extent of technology use, perceived as a proxy for participant exposure to a DBCI’s “active ingredients” or component behavior change techniques [10,11]. On the other hand, computer scientists have defined engagement as the subjective experience of “flow” or “immersion” that occurs during the human-computer interaction, characterized by focused attention, intrinsic interest, balance between challenge and skill, losing track of time and self-consciousness, and transportation to a “different place” [12,13]. After having conducted a systematic, integrative literature review of the psychology and computer science literatures [7] in addition to in-depth interviews with potential DBCI users, our interdisciplinary research team proposed the following working definition of engagement: “[Engagement with a DBCI is] a state-like construct which occurs each time a user interacts with a DBCI, with two behavioral (ie, amount and depth of use) and three experiential (ie, attention, interest and enjoyment) dimensions [14].”

We hence theorized that two behavioral (amount and depth of use) and three experiential (attention, interest, and enjoyment) dimensions are necessary and sufficient conditions for someone to be engaged with a DBCI. Although similar, engagement with DBCIs is thought to be conceptually distinct from both “flow” and pure technology usage. Although several measures of flow, immersion, and technology usage are available for use (for overviews, see [7,14,15]), an instrument that quantifies the intensity of behavioral and experiential engagement is lacking. For a quantitative scale of engagement to be useful for researchers, practitioners, and developers, it should be able to predict key variables of interest such as future engagement, knowledge acquisition, or intervention effectiveness. In addition, although a number of usage metrics derived from log-data are

typically used to capture the intensity of behavioral engagement [15-17], a validated measure of engagement, which captures both the experiential and behavioral dimensions of engagement and could be easily administered without the need to access and process the DBCI’s raw data, would be useful. The DBCI Engagement Scale was developed to fill this gap [14].

As part of the scale development process (described in detail in [14]), a pool of initial scale items was developed by the interdisciplinary research team in addition to two “best bets” for a short measure of engagement. Lay and expert respondents were then asked to classify the initial scale items into one of six categories (ie, *amount of use*, *depth of use*, *interest*, *attention*, *enjoyment*, plus an *unclassified* category) to examine the scale’s content validity. The first psychometric evaluation of the 10-item DBCI Engagement Scale was conducted in a sample of adult excessive drinkers who had voluntarily downloaded a freely available, evidence-informed app—*Drink Less*—for reducing their alcohol consumption [14]. Results indicated that the behavioral and experiential indicators of engagement may resolve to a single dimension. However, fewer than 5% of eligible users completed the scale during the study, and a sensitivity analysis indicated that the analytic sample was biased toward highly engaged users.

Studying engagement in real-world settings is notoriously difficult, as highly engaged users are more likely to respond to research surveys [18], potentially biasing results. Moreover, evidence suggests that motivation to change the target behavior is consistently associated with the frequency of behavioral engagement, such as the total number of logins [19,20]. Although motivation to change is a key predictor of engagement, it is neither a necessary nor a sufficient condition for someone to be engaged with a DBCI. For example, a user with low motivation to reduce their alcohol consumption might be intrigued by the design of a specific app, engage with its content, and subsequently become motivated to drink less. Therefore, to better study the dimensional structure of engagement, we considered it important to adjust for motivation to change in our analyses, thus separating the state of engagement from confounding motivations. This study aimed to evaluate the DBCI Engagement Scale in a sample of users recruited via an online research platform in order to address the following research questions:

1. What is the factor structure of the DBCI Engagement Scale? (construct validity)
2. Is the DBCI Engagement Scale internally reliable? (internal reliability)
3. Are total scale scores positively associated with objectively recorded amount of use and depth of use? (criterion validity)
4. Do total scale scores predict future behavioral engagement (ie, subsequent login), with and without adjustment for motivation to reduce alcohol consumption? (predictive validity)
5. Do two best bets for a short measure of engagement predict future behavioral engagement, with and without adjustment for motivation to reduce alcohol consumption? (predictive validity)
6. Does a model including the objectively recorded behavioral and the self-reported experiential indicators of engagement

account for more variance in future behavioral engagement (ie, subsequent login) compared with a model including only the objectively recorded behavioral indicators of engagement? (incremental validity)

7. Are total scale scores significantly associated with scores on the Flow State Scale? (divergent validity)

Methods

The preregistered study protocol can be found in the Open Science Framework [21]. Ethical approval was granted by University College London's Computer Science Departmental Research Ethics Chair (Project ID: UCLIC/1617/004/Staff Blandford HFDH).

Inclusion Criteria

Participants were eligible to take part in the study if they were aged ≥ 18 years; reported an Alcohol Use Disorders Identification Test (AUDIT) score ≥ 8 , indicating excessive alcohol consumption [22]; were residing in the United Kingdom; owned an iPhone capable of running iOS 8.0 software (ie, an iPhone 4S or later models); and were willing to download and explore an app for reducing alcohol consumption.

Sampling

Participants were recruited via the online research platform Prolific [23]. Individuals who take part in research via online platforms are primarily motivated by the financial incentives and are not necessarily interested in health behavior change [24]. Therefore, we expected that Prolific would enable us to recruit a sample of users with different levels of motivation to change. We did not, however, expect to recruit a sample that is

representative of the general population of excessive drinkers in the United Kingdom.

Sample Size

No formal sample size calculation was performed. Based on the psychometric literature, a 25:1 participant-to-item ratio (ie, a total of 250 participants) was considered desirable [25].

Measures

To determine eligibility and describe the sample, data were collected on age; sex (female or male); type of work (manual, nonmanual, or other); patterns of alcohol consumption, measured by the AUDIT [22]; motivation to reduce alcohol consumption, measured by the Motivation To Stop Scale [26-28]; country of residence (United Kingdom or other); iPhone ownership (yes or no); and willingness to download and explore an alcohol reduction app (yes or no).

For eligible participants who downloaded and explored the *Drink Less* app, data were collected on location during first use of the app (home, work, vehicle, public transport, restaurant/pub/café, other's home, other, or can't remember) and the 10-item DBCI Engagement Scale [14], which captures momentary behavioral (ie, amount, depth of use) and experiential (ie, attention, interest, enjoyment) engagement with DBCIs (Textbox 1). A detailed account of how the scale items were developed and tested in a group of experts and nonexperts can be found in a previous study [14].

Data were also collected on the below variables, which were used to test the scale's criterion, predictive, incremental, and divergent validity.

Textbox 1. The DBCI Engagement Scale.

Please answer the following questions with regard to your most recent use of the *Drink Less* app.

How strongly did you experience the following?

1. Interest
2. Intrigue
3. Focus
4. Inattention
5. Distraction
6. Enjoyment
7. Annoyance
8. Pleasure (*Measured on a 7-point scale with end- and middle-points anchored: “not at all,” “moderately,” and “extremely”*)
9. How much time (in minutes) do you roughly think that you spent on the app? (*Enter free text*)
10. Which of the app’s components do you remember visiting? (You can select multiple options)
 - a) Calendar (*coded by the researchers as “Self-Monitoring/Feedback”*)
 - b) Create and view goals (*coded as “Goal Setting”*)
 - c) What has and hasn’t worked (*coded as “Self-Monitoring/Feedback”*)
 - d) Create and view action plans (*coded as “Action Planning”*)
 - e) Your hangover and you (*coded as “Self-Monitoring/Feedback”*)
 - f) Review your drinking (*coded as “Normative Feedback”*)
 - g) Dashboard (*coded as “Self-Monitoring/Feedback”*)
 - h) Game (*coded as “Cognitive Bias Re-Training”*)
 - i) Drink + me (*coded as “Identity Change”*)
 - j) Useful information (*coded as “Other”*)
 - k) Other (*coded as “Other”*)
 - l) Can’t remember (*coded as “Other”*)

Indexed as a proportion of available modules (eg, 5/7×100=71.4).

Construct, Criterion, and Incremental Validity

A record of the number of app screens viewed was kept during participants’ first login session to derive the objectively recorded *amount of use* and *depth of use*, which were used to test the scale’s construct, criterion, and incremental validity. The screen view records were stored in an online database (*NodeChef*) and extracted using the free python library *pandas*. The variable *amount of use* was derived by calculating the time spent (in seconds) during participants’ first login session. The variable *depth of use* was derived by calculating the number of app components visited during participants’ first login session, indexed as a proportion (0-100) of the number of available components within the *Drink Less* app (ie, Goal Setting, Self-monitoring/Feedback, Action Planning, Normative Feedback, Cognitive Bias Re-Training, Identity Change, Other [29]).

Predictive Validity

A record of the number of app screens viewed was also kept over the next 14 days to derive the variable *subsequent login*, which was used to test the scale’s predictive validity. A subsequent login (yes vs no) was defined as a new screen view

following at least 30 minutes of inactivity [30]. As health apps are likely to be abandoned after users’ first login [31,32], the authors theorized that a useful measure of engagement should be able to distinguish between users who are likely to return to an app.

Two items that represented the authors’ *best bets* for a short measure of engagement (ie, “How much did you like the app?” and “How engaging was the app?”) were developed by the study team and used to test whether a short measure of engagement had superior predictive validity compared with the scale in its entirety. These items were not explicitly drawn from published self-report scales.

Divergent Validity

Two items from the Flow State Scale [33] were used to test the scale’s divergent validity. We selected two items that were previously found to load most strongly onto the general *flow* factor (ie, “When using *Drink Less*, the way time passed seemed to be different from normal,” “When using *Drink Less*, I was not worried about what others may have been thinking of me”). Although there is some overlap in the experiential indicators of the states of engagement and flow (ie, focused attention, interest), the study team theorized that users do not necessarily

experience *loss of time and consciousness* or *balance between challenge and skill* when engaging with a DBCI. Assessing whether users can be engaged without necessarily being in a state of flow was therefore considered a useful test of the scale's divergent validity. The Flow State Scale has previously been applied in the context of digital gaming [34].

Procedure

Interested participants were identified via the recruitment platform, Prolific, and received a compensation of £0.50 for completing the screening questionnaire, hosted by Qualtrics survey software (Provo, Utah). Eligible participants were invited via Prolific's internal email system and asked to download the *Drink Less* app from the Apple App Store. Participants were instructed to explore the *Drink Less* app in the way that they would explore any new app and were told that the researchers would monitor their app usage to assess what content they were interested in. For technical reasons, participants were told that they had to select the option *Interested in drinking less alcohol* when asked about why they were using the *Drink Less* app and to enable the push notifications. When clicking on the phone's home button after having finished exploring the app, participants received a push notification with a link to the study survey. Participants were subsequently asked to enter their Prolific identification number, which enabled the researchers to match participants' survey responses to their app screen views. Participants who initiated but did not complete the study survey (as indicated by their response status on Prolific's platform, which was either labelled "Timed out" or "Returned submission") were sent one reminder message. On completing the task, participants were paid £1.25.

Data Analysis

All analyses were conducted in SPSS version 20.0 (IBM Corporation, Armonk, New York). The assumptions for parametric tests were assessed (ie, normality of the distribution of residuals) and when violated, normalization was applied (ie, z-score normalization of positively skewed data). Descriptive statistics (eg, mean, range, variance) were calculated for each scale item and the criterion variables of interest to determine suitability for factor analysis.

Construct Validity

It was hypothesized that a five-factor solution (ie, *amount of use, depth of use, attention, interest, enjoyment*) would provide the best fit of the observed data [14]. A series of exploratory factor analyses (EFAs) using principal axis factoring estimation and oblique rotation was conducted. The inspection of Cattell's scree plots and the Kaiser criterion (ie, factors with eigenvalues >1) was used to determine the number of factors to retain [25]. First, we tested the fit of a solution including the self-reported items. This was compared with a solution including a combination of the self-reported indicators of experiential engagement and the objectively recorded indicators of

behavioral engagement (ie, objective *amount of use* and *depth of use*).

Internal Reliability

Internal consistency reliability was assessed by calculating the Cronbach alpha. A large coefficient (ie, $\geq .70$ or above) was interpreted as evidence of strong item covariance [35].

Criterion Validity

Criterion validity was assessed by calculating the Pearson correlation coefficient for the relationship between participants' automatically recorded app screen views from their first login (ie, objective *amount of use* and *depth of use*) with their self-reported *amount of use* and *depth of use* and their total scale scores.

Predictive Validity

The variable *subsequent login* was regressed onto participants' total scale scores, with and without adjustment for motivation to reduce alcohol consumption.

The variable *subsequent login* was also regressed onto each of the two *best bets* for a short measure of engagement (ie, "How engaging was the app?" and "How much did you like the app?"), with and without adjustment for motivation to reduce alcohol consumption.

Incremental Validity

Incremental validity was assessed in two steps. First, we assessed the variance accounted for in the variable *subsequent login* by the objectively recorded indicators of behavioral engagement. This was compared with the variance accounted for in the variable *subsequent login* after adding the self-reported indicators of experiential engagement to the objectively recorded indicators of behavioral engagement.

Divergent Validity

Divergent validity was assessed by calculating the Pearson correlation coefficient for the relationship between each of the two indicators of the state of flow from the Flow State Scale [33] and the overall measure of engagement.

Results

Participants

During the study period (31 days; July 23, 2018 to August 22, 2018), 401 participants completed the online screening survey, of which 266 were eligible to take part. Of these, 147 (55%) participants downloaded the *Drink Less* app and completed the task (Figure 1). Due to funding restrictions, we were unable to extend the recruitment beyond this time point. The desired target sample size of 250 participants was hence not achieved. Participants' demographic and drinking characteristics are reported in Table 1. We did not detect any significant differences between eligible participants who did and did not complete the task on the demographic characteristics assessed.

Figure 1. Participant flow chart. AUDIT: Alcohol Use Disorders Identification Test; ID: identification.

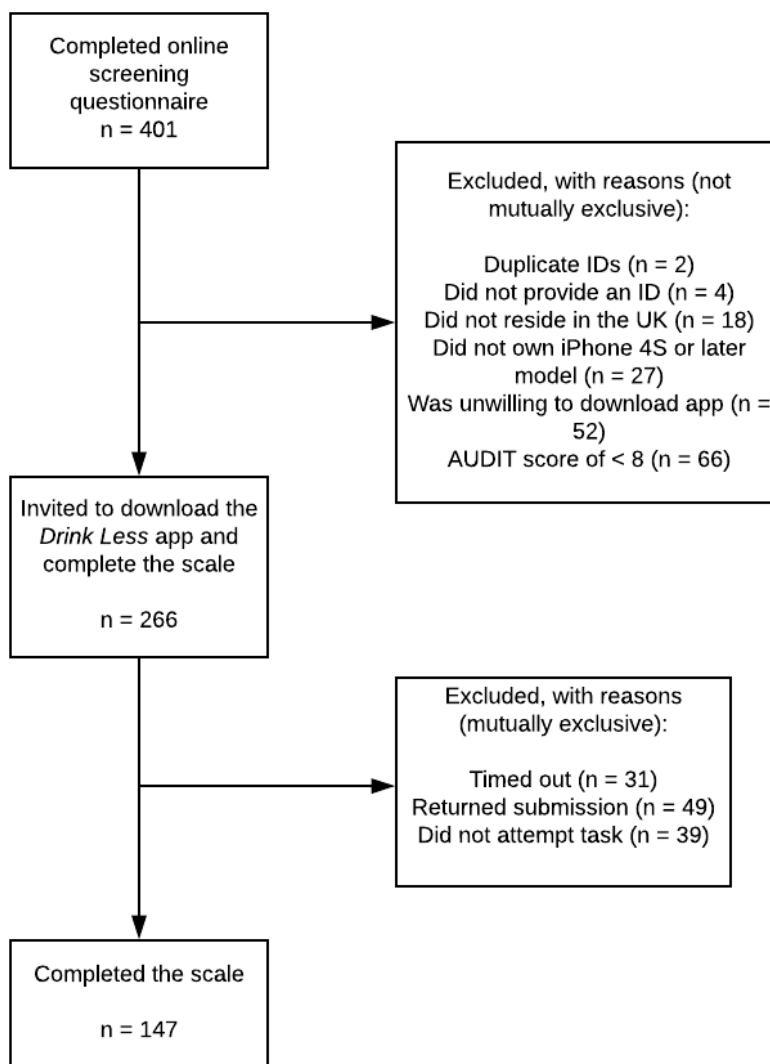


Table 1. Participants' demographic and drinking characteristics.

Demographic characteristics	Completed scale (n=147)	Eligible but did not complete scale (n=119)	<i>P</i> value ^a
Female gender, n (%)	97 (66)	71 (60)	.29
Type of work, n (%)			.57
Manual, n (%)	19 (13)	16 (13)	
Nonmanual, n (%)	89 (61)	78 (66)	
Other, n (%)	39 (27)	25 (21)	
Age (years), mean (SD)	34.4 (10.4)	36.6 (11.8)	.11
Drinking characteristics			
Motivation To Stop Scale, n (%)			.08
I don't want to cut down on drinking alcohol	14 (10)	26 (22)	
I think I should cut down on drinking alcohol but I don't really want to	43 (29)	25 (21)	
I want to cut down on drinking alcohol but I haven't thought about when	19 (13)	17 (14)	
I really want to cut down on drinking alcohol but I don't know when I will	17 (12)	11 (9)	
I want to cut down on drinking and hope to soon	23 (16)	17 (14)	
I really want to cut down on drinking alcohol and intend to in the next 3 months	11 (7)	4 (3)	
I really want to cut down on drinking alcohol and intend to in the next month	20 (14)	19 (16)	
Alcohol Use Disorders Identification Test, mean (SD)	15.4 (5.1)	14.2 (5.7)	.07

^aDifferences between groups were assessed using chi-square tests or *t* tests, as appropriate.

Descriptive Statistics

Descriptive statistics for the scale items are reported in [Table 2](#). The majority of participants completed the scale at home (118/147, 80.3%) or work (19/147, 12.9%). To account for the observed skewness, *z*-score normalization was applied to the 10-scale items and the two items used for testing the scale's criterion validity. Inter-item correlations of the normalized scale items are reported in [Table 3](#).

Scale Evaluation

Construct Validity

The Keiser-Meier Olkin Test of Sampling Adequacy (0.70) and the Bartlett Test of Sphericity ($P < .001$) indicated that data were suited for factor analysis. Three EFA solutions were tested to arrive at a best-fitting solution.

Table 2. Descriptive statistics for the scale items (N=147).

Statistic	Range	Mean (SD)	Variance	Skewness	Kurtosis
DBCI^a Engagement Scale Items					
1. "How strongly did you experience interest?"	2-7	5.30 (1.09)	1.18	-0.30	0.06
2. "How strongly did you experience intrigue?"	1-7	5.39 (1.27)	1.61	-0.85	0.50
3. "How strongly did you experience focus?"	2-7	5.31 (1.18)	1.40	-0.56	0.14
4. "How strongly did you experience inattention?" ^b	1-7	5.61 (1.33)	1.76	-1.24	1.47
5. "How strongly did you experience distraction?" ^b	1-7	5.47 (1.45)	2.10	-1.12	0.86
6. "How strongly did you experience enjoyment?"	1-7	4.46 (1.44)	2.07	-0.10	-0.48
7. "How strongly did you experience pleasure?"	1-7	3.56 (1.64)	2.67	0.36	-0.70
8. "How strongly did you experience annoyance?" ^b	1-7	5.59 (1.39)	1.93	-1.09	1.08
9. "Which of the app's components did you visit?"	14.29-100.00	58.70 (22.00)	484.01	-0.12	-0.67
10. "How much time do you roughly think that you spent on the app?" (seconds)	120-1200	520.82 (237.21)	56,267.82	0.93	0.96
Variables used to test the scale's construct, criterion, and incremental validity					
11. Objectively recorded depth of use	28.57-100.00	66.66 (20.50)	420.28	-0.23	-0.85
12. Objectively recorded amount of use (seconds)	95.00-3,571.00	409.45 (360.71)	130,116.72	5.13	40.34
Items used to test the scale's divergent validity					
13 "When using <i>Drink Less</i> , the way time passed seemed different from normal."	1-5	2.76 (0.79)	0.62	0.11	0.10
14. "When using <i>Drink Less</i> , I was not worried about what others may have been thinking about me."	1-5	3.34 (1.16)	1.35	-0.24	-1.11
Variables/items used to test the scale's predictive validity					
15. "How much did you like the app?"	1-7	5.14 (1.29)	1.66	-0.80	0.82
16. "How engaging was the app?"	1-7	5.20 (1.17)	1.37	-0.65	0.66
17. Subsequent login (yes vs no), n (%)	67 (46)	N/A ^c	N/A	N/A	N/A

^aDBCI: digital behavior change intervention.

^bValues were reverse scored prior to the calculation of descriptive statistics.

^cNot applicable.

Table 3. Inter-item correlation matrix (N=147).

DBCi Engagement Scale items	1 ^a	2 ^b	3 ^c	4 ^{d,e}	5 ^{e,f}	6 ^g	7 ^h	8 ^{e,i}	9 ^j	10 ^k	11 ^{l,m}	12 ^{m,n}
1. Interest	1											
<i>P</i> value	N/A ^o											
2. Intrigue	0.44	1										
<i>P</i> value	<.001	N/A										
3. Focus	0.65	0.46	1									
<i>P</i> value	<.001	<.001	N/A									
4. Inattention^e	0.18	0.10	0.31	1								
<i>P</i> value	.027	.21	<.001	N/A								
5. Distraction^e	0.18	0.12	0.28	0.43	1							
<i>P</i> value	.026	.16	.001	<.001	N/A							
6. Enjoyment	0.48	0.31	0.44	-0.05	-0.15	1						
<i>P</i> value	<.001	<.001	<.001	.59	.071	N/A						
7. Pleasure	0.19	0.09	0.15	-0.19	-0.24	0.54	1					
<i>P</i> value	.025	.30	.079	.019	.003	<.001	N/A					
8. Annoyance^e	0.28	0.15	0.37	0.27	0.24	0.29	0.12	1				
<i>P</i> value	.001	.07	<.001	.001	.004	<.001	.16	N/A				
9. Which of app's components	0.18	0.00	0.06	0.13	-0.03	0.19	0.19	.13	1			
<i>P</i> value	.028	.97	.469	.11	.75	.019	.02	.13	N/A			
10. How much time spent	0.10	0.10	-0.03	0.08	0.11	0.15	0.33	0.09	0.29	1		
<i>P</i> value	.23	.22	.681	.33	.18	.07	<.001	.31	<.001	N/A		
11. Objective depth of use^m	0.13	0.11	0.15	0.18	0.01	0.11	-0.01	0.24	0.51	0.16	1	
<i>P</i> value	.12	.18	.069	.03	.90	.20	.94	.003	<.001	.051	N/A	
12. Objective amount of use^m	0.31	0.18	0.28	0.16	0.06	0.25	0.00	0.19	0.10	0.10	0.52	1
<i>P</i> value	<.001	.03	.001	.047	.49	.002	.97	.02	.22	.23	<.001	N/A

^aInterest.

^bIntrigue.

^cFocus.

^dInattention.

^eValues were reverse scored prior to analysis.

^fDistraction.

^gEnjoyment.

^hPleasure.

ⁱAnnoyance.

^jWhich of the app's components.

^kHow much time spent.

^lObjective depth of use.

^mVariables used to test the scale's construct, criterion, and incremental validity.

ⁿObjective amount of use.

^oNot applicable.

Solution 1

An EFA with oblique rotation was conducted. The eigenvalues indicated that a three-factor solution, accounting for 61.2% of the variance, was most appropriate (Table 4). The loadings indicated that the second factor comprised two of the negatively worded indicators (ie, items 4 and 5). The third factor comprised the two behavioral indicators (ie, items 9 and 10) and one of the experiential indicators (ie, item 7), which made little theoretical sense [14]. The loading of item 8 (also a negatively worded item) onto factor 1 was modest. Therefore, the negatively worded items (ie, items 4, 5, and 8) and item 7 were discarded prior to conducting a second EFA.

Solution 2

A subsequent EFA with oblique rotation indicated that a two-factor solution accounted for 62.4% of the variance (Table 4). The experiential indicators loaded clearly onto factor 1, and the behavioral indicators loaded clearly onto factor 2, with no cross-loadings (ie, items that load at 0.32 or higher on two or

more factors) [25]. The two latent factors were labelled *Experiential Engagement* and *Behavioral Engagement*, respectively.

Solution 3

An EFA with oblique rotation using a combination of the self-reported experiential indicators (ie, items 1, 2, 3, and 6) and the automatically recorded behavioral indicators (ie, items 11 and 12) suggested a two-factor solution, which accounted for 65.7% of the variance. The experiential indicators loaded clearly onto factor 1, and the behavioral indicators loaded clearly onto factor 2 (Table 4).

Solution 2 was selected for use in the subsequent reliability and validity analyses, as it contained only the self-reported items and provided a similarly good fit of the data as Solution 3. A total scale score was calculated for each participant, with equal weight given to each of the retained items (ie, items 1, 2, 3, 6, 9, and 10).

Table 4. Factor loadings of the DBCI Engagement Scale in exploratory factor analyses.

Scale Items	Solution 1 ^a			Solution 2 ^b		Solution 3 ^c	
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 1	Factor 2
1. Interest	0.75 ^d	0.14	0.25	0.80 ^d	0.26	0.82 ^d	0.28
2. Intrigue	0.51 ^d	0.09	0.11	0.55 ^d	0.09	0.55 ^d	0.18
3. Focus	0.87 ^d	0.28	0.09	0.83 ^d	0.02	0.80 ^d	0.27
4. Inattention ^e	0.25	0.61 ^d	0.14	N/A ^f	N/A	N/A	N/A
5. Distraction ^e	0.21	0.68 ^d	0.06	N/A	N/A	N/A	N/A
6. Enjoyment	0.66 ^d	-0.35	0.43 ^d	0.57 ^d	0.31	0.57 ^d	0.23
7. Pleasure	0.31	-0.48	0.56 ^d	N/A	N/A	N/A	N/A
8. Annoyance ^e	0.41 ^d	0.23	0.25	N/A	N/A	N/A	N/A
9. Which of app's components	0.16	0.01	0.43 ^d	0.15	0.55	N/A	N/A
10. How much time spent	0.10	0.03	0.64 ^d	0.09	0.53	N/A	N/A
11. Objective depth of use	N/A	N/A	N/A	N/A	N/A	0.37	0.77 ^d
12. Objective amount of use	N/A	N/A	N/A	N/A	N/A	0.18	0.68 ^d

^aExploratory factor analysis with oblique rotation, including items 1-10.

^bExploratory factor analysis with oblique rotation, including items 1, 2, 3, 6, 9, and 10.

^cExploratory factor analysis with oblique rotation, including items 1, 2, 3, 6, 11, and 12.

^dValues with factor loadings ≥0.40.

^eValues were reverse scored prior to analysis.

^fNot applicable.

Internal Reliability

The internal consistency of the overall measure was 0.67, indicating moderate internal reliability [35]. The *Experiential Engagement* subscale had an internal consistency of 0.78, while the *Behavioral Engagement* subscale had an internal consistency of 0.45. Both subscales were significantly correlated with the measure overall ($r_{145}=0.90, P<.001$ and $r_{145}=0.56, P<.001$,

respectively). However, the subscales were not significantly correlated with each other ($r_{145}=0.15, P=.07$).

Criterion Validity

Total scale scores were significantly correlated with objectively recorded *depth of use* ($r_{145}=0.32, P<.001$) and objectively recorded *amount of use* ($r_{145}=0.33, P<.001$). Self-reported *depth of use* was significantly correlated with objectively recorded

depth of use ($r_{145}=0.51, P<.001$). Self-reported amount of use was not significantly correlated with objectively recorded amount of use ($r_{145}=0.10, P=.23$).

Predictive Validity

Results from the predictive validity analyses are presented in Table 5. In the unadjusted analysis, total scale scores were significantly associated with future behavioral engagement, ie, the variable subsequent login (odds ratio [OR]=1.15, 95% CI 1.05-1.27, $P=.01$). The association remained significant in the model adjusting for motivation to reduce alcohol consumption (adjusted OR [OR_{adj}]=1.14, 95% CI 1.03-1.27, $P=.01$).

As the two subscales (ie, Behavioral Engagement and Experiential Engagement) were not significantly correlated with

each other, an unplanned analysis was conducted to assess the independent association of each subscale with future behavioral engagement. In unadjusted and adjusted analyses, Experiential Engagement was significantly associated with future behavioral engagement (OR_{adj}=1.19, 95% CI 1.05-1.34, $P=.006$). In unadjusted and adjusted analyses, Behavioral Engagement was not significantly associated with future behavioral engagement (OR_{adj}=1.31, 95% CI 0.38-4.59, $P=.67$).

In unadjusted and adjusted analyses, asking users about how engaging they thought the app was did not significantly predict future behavioral engagement (OR_{adj}=1.34, 95% CI 0.98-1.84, $P=.07$). In unadjusted and adjusted analyses, asking users about how much they liked the app significantly predicted future behavioral engagement (OR_{adj}=1.38, 95% CI 1.03-1.84, $P=.03$).

Table 5. Unadjusted and adjusted odds ratios for the associations between the predictor variables and future behavioral engagement.

Predictor variables	Odds ratio (95% CI)	P value	Adjusted odds ratio ^a (95% CI)	P value
Total DBCI^b Engagement Scale score	1.15 (1.05-1.27)	.005	1.14 (1.03-1.27)	.009
Subscale 1 - Experiential Engagement	1.19 (1.06-1.34)	.004	1.19 (1.05-1.34)	.006
Subscale 2 - Behavioral Engagement	1.11 (0.90-1.36)	.34	1.08 (0.87-1.35)	.48
“How engaging was the app?”	1.28 (0.96-1.71)	.097	1.34 (0.98-1.84)	.07
“How much did you like the app?”	1.39 (1.05-1.83)	.02	1.38 (1.03-1.84)	.03

^aOdds ratios adjusted for motivation to reduce alcohol consumption.

^bDBCI: digital behavior change intervention.

Incremental Validity

Results from the incremental validity analyses are reported in Table 6. The automatically recorded behavioral indicators of engagement (ie, items 11 and 12; Model 1) accounted for 15.9%

of variance in the variable subsequent login. The automatically recorded behavioral indicators in combination with the self-reported experiential indicators of engagement (ie, items 1, 2, 3, and 6; Model 2) accounted for 21.1% of variance in the variable subsequent login.

Table 6. Odds ratios for the associations between the predictor variables and future behavioral engagement.

Models	Odds ratio (95% CI)	P value	Variance accounted for (%)
Model 1			15.9
Objectively recorded amount of use	3.46 (1.58-7.57)	.002	
Objectively recorded depth of use	0.91 (0.58-1.42)	.67	
Model 2			21.1
Objectively recorded amount of use	2.86 (1.25-6.55)	.013	
Objectively recorded depth of use	0.95 (0.60-1.50)	.82	
Interest	1.72 (1.03-2.85)	.04	
Focus	0.82 (0.50-1.35)	.44	
Enjoyment	0.93 (0.61-1.40)	.72	
Intrigue	1.17 (0.78-1.76)	.45	

Divergent Validity

Total scale scores were significantly correlated with the first (“When using Drink Less, the way time passed seemed different from normal”) but not the second (“When using Drink Less, I was not worried about what others may have been thinking about me”) indicator of flow ($r_{145}=0.25, P<.01$ and $r_{145}=-0.01,$

$P=.95,$ respectively). The two items tapping flow were not significantly correlated with one another in this sample ($r_{145}=-0.06, P=.47$).

Discussion

Principal Findings

The DBCI Engagement Scale was found to be underpinned by two, largely independent factors, which were labelled *Experiential Engagement* and *Behavioral Engagement*. The scale showed moderate internal reliability, but low divergent and criterion validity. Importantly, the behavioral subscale may not be a valid indicator of behavioral engagement. Total scale scores were weakly associated with future behavioral engagement (ie, the variable *subsequent login*), as were the experiential subscale and one of the *best bets* for a short measure of engagement (ie, asking participants about how much they liked the app). The behavioral subscale was not independently associated with future behavioral engagement. In addition, a model including the self-reported experiential and objectively recorded behavioral indicators of engagement (as compared with a model including only the objectively recorded behavioral indicators) accounted for a larger proportion of variance in future behavioral engagement. These findings are at odds with those from the first evaluation of the DBCI Engagement Scale, in which the scale was found to be underpinned by a single factor [14]. However, these differences may at least partly be accounted for by the small sample size in this study.

Construct Validity

The finding that the *Experiential Engagement* and *Behavioral Engagement* subscales were not significantly correlated with each other in this study lends support to the argument that users can spend time on a DBCI without necessarily being interested in or paying attention to its content, and vice versa [14]. However, this finding also gives rise to the question of whether experiential and behavioral engagement are part of the same higher-order construct.

The finding that participants' total scale scores were weakly associated with future behavioral engagement even when adjusting for motivation to reduce alcohol consumption serves as initial evidence that the state of engagement with a DBCI is conceptually distinct from motivation to change the target behavior.

Incremental and Predictive Validity

The results from the incremental validity analyses suggest that behavioral and experiential indicators in tandem have superior predictive power compared with the behavioral indicators alone. However, the finding that the experiential, but not the behavioral, subscale was independently associated with future behavioral engagement can be interpreted to suggest that the experiential indicators (particularly users' interest) were driving the association between initial and future engagement. A potential explanation for these findings is that more intensive engagement during the first login session might have made users' memory of the app more salient, which might have made them more likely to remember to return to the app. As one of the short measures of engagement (ie, the item asking about how much users liked the app) was also found to predict future engagement, it is possible that not only salience of the app, but a salient memory of liking the app, is important for future

engagement. It is unclear why the first, but not the second, short measure of engagement had significant predictive power; the word *liking* might be easier to interpret than the word *engaging*. The potential mechanisms underlying the relationship between initial experiential and behavioral engagement, and future behavioral engagement (ie, the variable *subsequent login*) should be explored further using experience sampling techniques in the first few hours following initial app engagement; this involves repeated measurements of psychological processes in real time, in users' natural environments [36].

These results also beg the question as to whether future behavioral engagement is the most appropriate criterion variable to test an engagement scale against. For example, knowledge retention or skill acquisition may be more theoretically sound, as suggested by the Elaboration Likelihood Model of Persuasion (ELMP) [37]. The ELMP argues that deep information processing occurs when an individual pays attention to (or engages with) a health message, which leads to increased knowledge retention. It is plausible that initial behavioral and experiential engagement have superior predictive power compared with behavioral engagement when used to predict knowledge retention. In addition, it would be useful to assess whether the new measure of moment-to-moment (or state-like) engagement is able to predict intervention effectiveness at a later time point.

Criterion Validity

The finding that the self-reported and objectively recorded indicators of *amount of use* were not significantly correlated in this sample suggests that the DBCI Engagement Scale may not be a valid indicator of behavioral engagement. However, although the *amount of use* (ie, time spent in minutes or seconds) is typically used as a gold standard or ground truth of behavioral engagement, our results showed that objectively recorded *amount of use* was significantly correlated with many of the experiential indicators (eg, interest, intrigue). Although the exploratory factor analyses did not indicate that *amount of use* loads onto the same factor as the experiential indicators of engagement, the observed pattern of correlations leads us to question whether time spent on a DBCI is deserving of its ground truth status. There is, hence, a need for future research to investigate the source of the discrepancy between self-reported and objectively recorded indicators of *amount of use*.

Divergent Validity

In line with the first study evaluating the scale, this study did not provide evidence that the DBCI Engagement Scale diverges from the Flow State Scale. There is conceptual overlap between engagement with DBCIs and the dimension of flow that is labelled *losing track of time*. It should be noted that the proposed definition of engagement was, in part, developed based on the concept of flow [14]. It may hence be more fruitful to assess the scale's divergent validity using a more conceptually distinct measure in the future. The lack of evidence that the DBCI Engagement Scale diverges from the Flow State Scale may also serve as a plausible explanation for why participants' self-reported *amount of use* was not significantly correlated with their objectively recorded *amount of use*; they may have

lost track of time when engaging with the *Drink Less* app. This finding suggests that self-reported and objectively recorded indicators of time spent on a DBCI may tap different constructs; future research is required to examine which of these is more strongly related to key outcomes of interest.

Limitations

This study was limited because it did not achieve the desired sample size of 250 participants. As Prolific is a novel platform with a small proportion of individuals meeting the study eligibility criteria (ie, drinking alcohol excessively, willing to download an alcohol reduction app, owning an iPhone), the extant participant pool was exhausted after screening just over 400 participants. Although the participant-to-item ratio is considered key in determining the minimum necessary sample size for conducting factor analyses, findings from simulation studies indicate that other factors, including the number of items per factor and the level of communality between items, also influence sample size requirements [38]. Given the limited participant-to-item ratio and the small number of items per factor in this study, the two-factor solution should be interpreted with caution and merits replication in a larger sample in future research. A second limitation is that market research indicates that iOS users are, on average, more affluent than Android users [39]. As the *Drink Less* app is currently available for iOS users only, our findings may not be generalizable to Android users.

Studies conducted via Prolific that involve an initial screening study followed by inviting eligible participants to complete the actual study tend to have attrition rates of approximately 20%-25%, and not 45% [40]. It is therefore likely that there were systematic differences between eligible participants who completed the task and those who did not. For example, the small financial reward may not have been perceived as worth the effort of downloading an app. Indeed, a study assessing the demographic and psychological characteristics of participants who regularly complete research tasks via Amazon's Mechanical Turk online platform (which is similar to Prolific) found that the majority of surveyed participants reported that earning money was a key motivator for taking part [24]. It should also be noted that the financial incentive may have interfered with participants' naturalistic engagement, thus limiting the generalizability of the findings. Previous research has found that money can be an important motivator in DBCI research and increase response rates in longitudinal studies [41].

We did not want to overburden users; hence, we did not assess key trait-like variables that may have influenced users' scale scores. For example, it would have been useful to attempt to partial out the variance accounted for by users' personality traits, such as those specified in the Big Five model of personality

[42], to ensure that the DBCI Engagement Scale is detecting something beyond high conscientiousness or low neuroticism.

The adjustment for participants' motivation to reduce their alcohol consumption should have increased the item covariance on the DBCI Engagement Scale and is hence considered a study strength. It should, however, be noted that participants' motivation may have interacted with their engagement levels. Hence, despite the adjustment for participants' motivation to change, the scale scores may not fully represent participants' "true" engagement scores.

Finally, the decision to use Google's cutoff (ie, 30 minutes of inactivity) to identify whether users had made a subsequent login is, to our best knowledge, not grounded in evidence about session length. Future research should explore whether this constitutes a useful heuristic for identifying new DBCI sessions using both quantitative and qualitative methods.

Avenues for Future Research

Due to the observed nonnormal distributions of the scale items that jointly form the DBCI Engagement Scale, a decision was made to use z-score normalization. Consequently, total scores on the DBCI Engagement Scale are only meaningful in relation to the average intensity of experiential and behavioral engagement that a particular DBCI generates. This may facilitate attempts to develop cutoffs for "high" and "low" engagers across DBCIs, irrespective of their specific parameters (eg, the number and length of intervention components). For example, users with scores that fall within a particular range of SDs above or below the mean might usefully be classified as "high" or "low" engagers, and these patterns may replicate across DBCIs. The question of whether the mean and spread of engagement scores replicate across DBCIs merits exploration by evaluating the DBCI Engagement Scale across different kinds of DBCIs (eg, websites or apps for smoking cessation or physical activity).

The finding that initial experiential engagement (or liking of the app) was independently associated with future behavioral engagement suggests that intervention developers should think carefully about how to make their DBCIs appealing on first use. The DBCI Engagement Scale may be useful during the iterative design process, comparing users' experiences of differently designed graphical user interfaces.

Conclusions

The DBCI Engagement Scale assesses behavioral and experiential aspects of engagement. The behavioral subscale may not be a valid indicator of behavioral engagement. The experiential subscale can predict subsequent behavioral engagement with an app for reducing alcohol consumption. Further refinements and validation of the scale in larger samples and across different DBCIs are needed.

Acknowledgments

OP is funded by a PhD studentship from Bupa under its partnership with UCL. CG receives salary support from Cancer Research UK (C1417/A22962). SM is funded by Cancer Research UK and the NIHR School for Public Health Research. We gratefully acknowledge all the funding received.

Authors' Contributions

OP, JL, CG, AB, RW, and SM designed the study. OP collected the data, conducted the statistical analyses, and wrote the first draft of the manuscript. All authors have contributed to the final version of the manuscript and agree with its submission to JMIR.

Conflicts of Interest

OP, CG, AB, and SM have no conflicts of interest to declare. RW undertakes research and consultancy and receives fees for speaking from companies that develop and manufacture smoking cessation medications. JL is an employee at Prolific.

References

1. Donkin L, Christensen H, Naismith S, Neal B, Hickie I, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. *J Med Internet Res* 2011 Aug 05;13(3):e52 [FREE Full text] [doi: [10.2196/jmir.1772](https://doi.org/10.2196/jmir.1772)] [Medline: [21821503](https://pubmed.ncbi.nlm.nih.gov/21821503/)]
2. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. *J Med Internet Res* 2017 Jun 29;19(6):e232 [FREE Full text] [doi: [10.2196/jmir.7126](https://doi.org/10.2196/jmir.7126)] [Medline: [28663162](https://pubmed.ncbi.nlm.nih.gov/28663162/)]
3. Kelders S, Kok R, Ossebaard H, van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. *J Med Internet Res* 2012 Nov 14;14(6):e152 [FREE Full text] [doi: [10.2196/jmir.2104](https://doi.org/10.2196/jmir.2104)] [Medline: [23151820](https://pubmed.ncbi.nlm.nih.gov/23151820/)]
4. Cobb NK, Graham AL, Bock BC, Papandonatos G, Abrams DB. Initial evaluation of a real-world Internet smoking cessation system. *Nicotine Tob Res* 2005 Apr;7(2):207-216 [FREE Full text] [doi: [10.1080/14622200500055319](https://doi.org/10.1080/14622200500055319)] [Medline: [16036277](https://pubmed.ncbi.nlm.nih.gov/16036277/)]
5. Alexander GL, McClure JB, Calvi JH, Divine GW, Stopponi MA, Rolnick SJ, et al. A randomized clinical trial evaluating online interventions to improve fruit and vegetable consumption. *Am J Public Health* 2010 Feb;100(2):319-326 [FREE Full text] [doi: [10.2105/AJPH.2008.154468](https://doi.org/10.2105/AJPH.2008.154468)] [Medline: [20019315](https://pubmed.ncbi.nlm.nih.gov/20019315/)]
6. Yardley L, Spring B, Riper H, Morrison L, Crane D, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. *Am J Prev Med* 2016 Nov;51(5):833-842 [FREE Full text] [doi: [10.1016/j.amepre.2016.06.015](https://doi.org/10.1016/j.amepre.2016.06.015)] [Medline: [27745683](https://pubmed.ncbi.nlm.nih.gov/27745683/)]
7. 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 Jun;7(2):254-267 [FREE Full text] [doi: [10.1007/s13142-016-0453-1](https://doi.org/10.1007/s13142-016-0453-1)] [Medline: [27966189](https://pubmed.ncbi.nlm.nih.gov/27966189/)]
8. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. *J Med Internet Res* 2017 Dec 06;19(12):e402 [FREE Full text] [doi: [10.2196/jmir.8578](https://doi.org/10.2196/jmir.8578)] [Medline: [29212630](https://pubmed.ncbi.nlm.nih.gov/29212630/)]
9. Milward J, Drummond C, Fincham-Campbell S, Deluca P. What makes online substance-use interventions engaging? A systematic review and narrative synthesis. *Digit Health* 2018;4:2055207617743354 [FREE Full text] [doi: [10.1177/2055207617743354](https://doi.org/10.1177/2055207617743354)] [Medline: [29942622](https://pubmed.ncbi.nlm.nih.gov/29942622/)]
10. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med* 2013 Aug;46(1):81-95. [doi: [10.1007/s12160-013-9486-6](https://doi.org/10.1007/s12160-013-9486-6)] [Medline: [23512568](https://pubmed.ncbi.nlm.nih.gov/23512568/)]
11. Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, et al. Enhancing Treatment Fidelity in Health Behavior Change Studies: Best Practices and Recommendations From the NIH Behavior Change Consortium. *Health Psychology* 2004;23(5):443-451. [doi: [10.1037/0278-6133.23.5.443](https://doi.org/10.1037/0278-6133.23.5.443)]
12. Csikszentmihalyi M. *Flow: The psychology of optimal performance*. New York: Cambridge University Press; 1990.
13. Brown E, Cairns P. ACM Digital Library. 2004. A grounded investigation of game immersion. URL: <https://dl.acm.org/citation.cfm?doi=985921.986048> [accessed 2019-11-12]
14. 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* 2019 Mar 30. [doi: [10.1093/tbm/ibz039](https://doi.org/10.1093/tbm/ibz039)] [Medline: [30927357](https://pubmed.ncbi.nlm.nih.gov/30927357/)]
15. 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 Med Internet Res* 2018 Nov 16;20(11):e292. [doi: [10.2196/jmir.9397](https://doi.org/10.2196/jmir.9397)]
16. 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 Jan 18;7(1):e11941. [doi: [10.2196/11941](https://doi.org/10.2196/11941)]
17. Miller S, Ainsworth B, Yardley L, Milton A, Weal M, Smith P, et al. A Framework for Analyzing and Measuring Usage and Engagement Data (AMUsED) in Digital Interventions: Viewpoint. *J Med Internet Res* 2019 Feb 15;21(2):e10966. [doi: [10.2196/10966](https://doi.org/10.2196/10966)]

18. Murray E, White IR, Varaganam M, Godfrey C, Khadjesari Z, McCambridge J. Attrition revisited: adherence and retention in a web-based alcohol trial. *J Med Internet Res* 2013;15(8):e162 [FREE Full text] [doi: [10.2196/jmir.2336](https://doi.org/10.2196/jmir.2336)] [Medline: [2396958](https://pubmed.ncbi.nlm.nih.gov/2396958/)]
19. Postel M, de Haan HA, ter Huurne ED, van der Palen J, Becker E, de Jong CAJ. Attrition in web-based treatment for problem drinkers. *J Med Internet Res* 2011 Dec 27;13(4):e117 [FREE Full text] [doi: [10.2196/jmir.1811](https://doi.org/10.2196/jmir.1811)] [Medline: [22201703](https://pubmed.ncbi.nlm.nih.gov/22201703/)]
20. Radtke T, Ostergaard M, Cooke R, Scholz U. Web-Based Alcohol Intervention: Study of Systematic Attrition of Heavy Drinkers. *J Med Internet Res* 2017 Jun 28;19(6):e217 [FREE Full text] [doi: [10.2196/jmir.6780](https://doi.org/10.2196/jmir.6780)] [Medline: [28659251](https://pubmed.ncbi.nlm.nih.gov/28659251/)]
21. Perski O, Lumsden J, Garnett C, Blandford A, West R, Michie S. Second Evaluation of the DBCI Engagement Scale. *Open Sci Framew* 2019 [FREE Full text]
22. Babor T, Higgins-Biddle J, Saunders J, Monteiro M. The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Care. 2nd ed. Geneva: World Health Organization; 2001.
23. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology* 2017 May;70:153-163. [doi: [10.1016/j.jesp.2017.01.006](https://doi.org/10.1016/j.jesp.2017.01.006)]
24. Paolacci G, Chandler J, Ipeirotis P. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 2010;5(5):411-419 [FREE Full text] [Medline: [1011330798](https://pubmed.ncbi.nlm.nih.gov/1011330798/)]
25. Costello A, Osborne JW. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation* 2005;10:173-178.
26. Kotz D, Brown J, West R. Predictive validity of the Motivation To Stop Scale (MTSS): a single-item measure of motivation to stop smoking. *Drug Alcohol Depend* 2013 Feb 01;128(1-2):15-19 [FREE Full text] [doi: [10.1016/j.drugalcdep.2012.07.012](https://doi.org/10.1016/j.drugalcdep.2012.07.012)] [Medline: [22943961](https://pubmed.ncbi.nlm.nih.gov/22943961/)]
27. Hummel K, Brown J, Willemsen MC, West R, Kotz D. External validation of the Motivation To Stop Scale (MTSS): findings from the International Tobacco Control (ITC) Netherlands Survey. *Eur J Public Health* 2017 Feb 01;27(1):129-134. [doi: [10.1093/eurpub/ckw105](https://doi.org/10.1093/eurpub/ckw105)] [Medline: [28177479](https://pubmed.ncbi.nlm.nih.gov/28177479/)]
28. de Vocht F, Brown J, Beard E, Angus C, Brennan A, Michie S, et al. Temporal patterns of alcohol consumption and attempts to reduce alcohol intake in England. *BMC Public Health* 2016 Sep 01;16:917 [FREE Full text] [doi: [10.1186/s12889-016-3542-7](https://doi.org/10.1186/s12889-016-3542-7)] [Medline: [27585991](https://pubmed.ncbi.nlm.nih.gov/27585991/)]
29. Crane D, Garnett C, Michie S, West R, Brown J. A smartphone app to reduce excessive alcohol consumption: Identifying the effectiveness of intervention components in a factorial randomised control trial. *Sci Rep* 2018 Mar 12;8(1):4384 [FREE Full text] [doi: [10.1038/s41598-018-22420-8](https://doi.org/10.1038/s41598-018-22420-8)] [Medline: [29531280](https://pubmed.ncbi.nlm.nih.gov/29531280/)]
30. Google Analytics Help. 2017. How a web session is defined in Analytics. URL: <https://support.google.com/analytics/answer/2731565> [accessed 2018-02-06]
31. Braze Magazine. Spring 2016 Mobile Customer Retention Report. 2016. URL: <https://www.braze.com/blog/app-customer-retention-spring-2016-report/> [accessed 2019-09-09]
32. CISION PRWeb. 2015. Motivating Patients to Use Smartphone Health Apps. URL: <http://www.prweb.com/releases/2011/04/prweb5268884.htm> [accessed 2015-08-10]
33. Jackson S, Marsh H. Development and validation of a scale to measure optimal experience: The Flow State Scale. *J Sport Exerc Psychol* 1996;18:35 [FREE Full text]
34. Wiebe E, Lamb A, Hardy M, Sharek D. Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. *Comput Human Behav* 2014;32:132. [doi: [10.1016/j.chb.2013.12.001](https://doi.org/10.1016/j.chb.2013.12.001)]
35. Hinkin T. A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires. *Organ Res Methods* 1998;1(1):121. [doi: [10.1177/109442819800100106](https://doi.org/10.1177/109442819800100106)]
36. Stone A, Shiffman S. Ecological Momentary Assessment (Ema) in Behavioral Medicine. *Annals of Behavioral Medicine* 1994;16(3):199-202. [doi: [10.1093/abm/16.3.199](https://doi.org/10.1093/abm/16.3.199)]
37. Petty R, Cacioppo J. The Elaboration Likelihood Model of Persuasion. *Adv Exp Soc Psychol* 1986;19:205. [doi: [10.1016/S0065-2601\(08\)60214-2](https://doi.org/10.1016/S0065-2601(08)60214-2)]
38. Mundfrom DJ, Shaw DG, Ke TL. Minimum Sample Size Recommendations for Conducting Factor Analyses. *International Journal of Testing* 2005;5(2):159-168. [doi: [10.1207/s15327574ijt0502_4](https://doi.org/10.1207/s15327574ijt0502_4)]
39. Berg M. Statista. 2014. iPhone Users Earn More. URL: <https://www.statista.com/chart/2638/top-line-platform-stats-for-app-usage-in-the-us/> [accessed 2019-01-28]
40. Palan S, Schitter C. Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 2018 Mar;17:22-27 [FREE Full text] [doi: [10.1016/j.jbef.2017.12.004](https://doi.org/10.1016/j.jbef.2017.12.004)]
41. Khadjesari Z, Murray E, Kalaitzaki E, White I, McCambridge J, Thompson S, et al. Impact and costs of incentives to reduce attrition in online trials: two randomized controlled trials. *J Med Internet Res* 2011 Mar 02;13(1):e26 [FREE Full text] [doi: [10.2196/jmir.1523](https://doi.org/10.2196/jmir.1523)] [Medline: [21371988](https://pubmed.ncbi.nlm.nih.gov/21371988/)]
42. John OP, Donahue EM, Kentle RL. The Big Five Inventory. 1991. URL: http://www.sjdm.org/dmidi/Big_Five_Inventory.html [accessed 2019-09-09]

Abbreviations**AUDIT:** Alcohol Use Disorders Identification Test**DBCI:** digital behavior change intervention

Edited by G Eysenbach; submitted 09.09.19; peer-reviewed by S Miller, J Thrul; comments to author 17.10.19; revised version received 30.10.19; accepted 11.11.19; published 17.11.19

*Please cite as:**Perski O, Lumsden J, Garnett C, Blandford A, West R, Michie S**Assessing the Psychometric Properties of the Digital Behavior Change Intervention Engagement Scale in Users of an App for Reducing Alcohol Consumption: Evaluation Study**J Med Internet Res 2019;21(11):e16197**URL: <http://www.jmir.org/2019/11/e16197/>**doi: [10.2196/16197](https://doi.org/10.2196/16197)**PMID:*

©Olga Perski, Jim Lumsden, Claire Garnett, Ann Blandford, Robert West, Susan Michie. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 20.11.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.