

Original Paper

# Examining Predictors of Real-World User Engagement with Self-Guided eHealth Interventions: Analysis of Mobile Apps and Websites Using a Novel Dataset

Amit Baumel<sup>1</sup>, PhD; John M Kane<sup>2</sup>, MD

<sup>1</sup>Department of Community Mental Health, University of Haifa, Haifa, Israel

<sup>2</sup>Zucker Hillside Hospital, Glen Oaks, NY, United States

**Corresponding Author:**

Amit Baumel, PhD

Department of Community Mental Health

University of Haifa

Abba Khoushy Ave 199

Haifa, 3498838

Israel

Phone: 972 48288496

Email: [abaumel@univ.haifa.ac.il](mailto:abaumel@univ.haifa.ac.il)

## Abstract

**Background:** The literature suggests that the product design of self-guided electronic health (eHealth) interventions impacts user engagement. Traditional trial settings, however, do not enable the examination of these relationships in real-world use.

**Objective:** This study aimed to examine whether the qualities of product design, research evidence, and publicly available data predict real-world user engagement with mobile and Web-based self-guided eHealth interventions.

**Methods:** This analysis included self-guided mobile and Web-based eHealth interventions available to the public—with their qualities assessed using the Enlight suite of scales. Scales included Usability, Visual Design, User Engagement, Content, Therapeutic Persuasiveness, Therapeutic Alliance, Credibility, and Research Evidence. Behavioral data on real-world usage were obtained from a panel that provides aggregated nonpersonal information on user engagement with websites and mobile apps, based on a time window of 18 months that was set between November 1, 2016 and April 30, 2018. Real-world user engagement variables included average usage time (for both mobile apps and websites) and mobile app user retention 30 days after download.

**Results:** The analysis included 52 mobile apps (downloads median 38,600; interquartile range [IQR] 116,000) and 32 websites (monthly unique visitors median 5689; IQR 30,038). Results point to moderate correlations between Therapeutic Persuasiveness, Therapeutic Alliance, and the 3 user engagement variables ( $.31 \leq r_s \leq .51$ ;  $P_s \leq .03$ ). Visual Design, User Engagement, and Content demonstrated similar degrees of correlation with mobile app engagement variables ( $.25 \leq r_s \leq .49$ ;  $P_s \leq .04$ ) but not with average usage time of Web-based interventions. Positive correlations were also found between the number of reviews on Google Play and average app usage time ( $r = .58$ ;  $P < .001$ ) and user retention after 30 days ( $r = .23$ ;  $P = .049$ ). Although several product quality ratings were positively correlated with research evidence, the latter was not significantly correlated with real-world user engagement. Hierarchical stepwise regression analysis revealed that either Therapeutic Persuasiveness or Therapeutic Alliance explained 15% to 26% of user engagement variance. Data on Google Play (number of reviews) explained 15% of the variance of mobile app usage time above Enlight ratings; however, publicly available data did not significantly contribute to explaining the variance of the other 2 user-engagement variables.

**Conclusions:** Results indicate that the qualities of product design predict real-world user engagement with eHealth interventions. The use of real-world behavioral datasets is a novel way to learn about user behaviors, creating new avenues for eHealth intervention research.

(*J Med Internet Res* 2018;20(12):e11491) doi: [10.2196/11491](https://doi.org/10.2196/11491)

**KEYWORDS**

eHealth; mHealth; user engagement; user experience; therapeutic alliance; persuasive design; behavior change

## Introduction

### Background

Self-guided electronic health (eHealth) interventions have the potential to increase access to evidence-based care, while reducing the costs associated with service uptake [1,2]. The impact of these interventions, however, is limited by the ability to engage users in therapeutic activities and to support users' adherence to the therapeutic process [3,4]. As eHealth interventions require individuals to engage with self-care outside of traditional health care settings [5-7], individuals' engagement must compete with other events in their daily lives and fluctuating motivation to engage in effortful behavior [8]. As a result, user engagement with mobile apps and websites across the behavior change spectrum is low in the absence of human support [9].

There is a body of literature examining how intervention design may facilitate engagement and behavior change [10-14]. For example, systematic reviews have found relationships between the incorporation of therapeutic persuasiveness (ie, persuasive design and behavior change techniques) into the eHealth intervention and user adherence [15] and the intervention's efficacy [16]. Studies have also shown that the incorporation of conversational agents within self-guided eHealth interventions impacts user engagement [17,18], suggesting that conversational agents enhance the relational factors within the program—factors that are part of the therapeutic alliance fostered between the user and the program [19-21]. Although these studies provide convincing evidence linking product design to user engagement, the understanding of these relationships is still limited by our ability to examine user engagement in the real world, while comparing large numbers of products within the same study.

### Comparing a Large Number of Products

Certain methodologies enable the cost-effective comparison of a large number of interventions. The Multiphase Optimization Strategy, for example, offers a paradigm to incorporate a randomized experimentation in a way that directly compares many different intervention components to identify the group of active components leading to the best outcome [22]. When it comes to product designs that are largely different in their functionalities, however, to control all the active components, the different functions would have to be available within the utilized digital platform—a process that would be highly expensive given the average development price of a single health app [23].

A novel way to compare large numbers of digital products is to utilize datasets that record user behaviors on a large number of websites and mobile apps. This approach can enable the documentation of variance in user engagement across a wide range of product designs using the same analytical framework. The big data commonly generated and stored by digital platforms can be used to learn about user behaviors in order to refine conceptual models or theories, and to understand processes related to eHealth interventions [24]. Additionally, using commonly generated data enables us to record real-world user behaviors outside of study settings that involve interactions

with research staff—interactions that might impact user engagement [25,26].

### Identifying the Different Quality Aspects of Product Design

To utilize commonly generated data and to understand their relationship with eHealth product design, there is also a need to identify the different aspects of product design in a reliable way. The literature suggests several approaches to evaluating eHealth product design (see the study by BinDhim et al [27] for a review). These include using a predefined list of what the app should contain, assessing the inclusion of evidence-based content, assessing the usability of predefined app functions, using consumer reviews and ratings, and utilizing criteria-based rating scales. Criteria-based rating scales rest on a heuristic evaluation approach—a method that has been broadly researched and used for assessing eHealth and technology products [28-30]. Heuristics are broad principles of product design that can be inspected by evaluators before empirical testing. The advantage of heuristic evaluation is that it enables the cost-efficient identification of design problems without the need for a predefined list of features that may exclude the users' experience of utilizing the product [29,31,32]. Criteria-based rating scales are based on core concepts, each comprising different heuristics, which are used by trained raters to objectively examine and score the quality of eHealth programs [12,33-35]. In the absence of such clearly defined rating systems, scoring tends to be less reliable [36].

Enlight is a suite of criteria-based rating scales that was used in this study. It covers 11 different quality constructs and checklists (eg, Usability and Credibility) that were produced by trained raters. Enlightenment is the first suite of measures to incorporate behavior change, persuasive design, and therapeutic alliance concepts—concepts that have been found to affect a program's therapeutic potential [15,16]. As the tool shows high inter-rater reliability scores at the construct level, it enables us to use it to examine the relationships between different aspects of product design and metrics of user engagement.

Within this context, Baumel and Yom-Tov conducted a preliminary investigation, examining the correlations between 6 quality aspects of product design and real-world user engagement with 30 Web-based programs [37]. Real-world user engagement was based on a proprietary dataset of anonymized logs from consenting users of a Microsoft Internet Explorer add-on. The quality scores were generated using Enlightenment [21]. In the preliminary study, it was found that product quality ratings predicted which Web-based interventions were more engaging and in particular that Therapeutic Persuasiveness was the most robust predictor of user adherence (ie, duration of use, number of unique sessions;  $40 \leq r_s \leq .58$ ;  $P_s \leq .005$ ) [37].

Although these findings were novel in terms of the methods applied, they were limited in several aspects. First, the dataset only included Web-based interventions, whereas the literature suggests that there has been a massive increase in mobile phone ownership and mobile health app usage in recent years [38]. Second, as the analysis was based on a small sample of interventions, the question arises as to whether the same pattern of results will be found with a larger dataset, enabling

identification of correlations with other aspects of product design. Third, the study did not incorporate important metrics that relate to research evidence, a product's credibility, and publicly available data on user acceptance (eg, user ratings on app stores). Examining the relationship between research evidence and real-world uptake is key in light of the notion that efficacy trials largely emphasize internal validity over real-world issues, such as the technological environment, implementation, and sustainability, and thus may not provide the needed validation [26]. Furthermore, to the best of our knowledge, there has been no study linking the credibility of the source that developed the program or user ratings on app stores and real-world user engagement with eHealth programs.

### This Study

The aims of this study were, therefore, to fill this gap in the literature by (1) examining whether different qualities of product design predict real-world user engagement with both mobile- and Web-based self-guided eHealth interventions; (2) exploring the associations between scale items, data, and real-world user engagement; (3) examining whether research evidence and product credibility metrics are associated with real-world user engagement; (4) examining the associations between publicly available data on program acceptance (eg, star ratings on app stores) and real-world usage and whether these data enhance the prediction of user engagement above expert ratings; and (5) establishing Enlight's validity in predicting user engagement with eHealth interventions based on a large and independent dataset of user behaviors.

## Methods

### Selection of Interventions

We screened for eligibility all eHealth programs that were assessed based on Enlight suite of scales between September 2016 and December 2017—during the scale development phase and afterward as part of a nonprofit project aimed at evaluating the quality of eHealth programs [39]. The clinical aims of the selected programs were broad, spanning the behavioral health domain. These programs could be grouped into those targeting mental health (eg, depression, anxiety, and well-being) and those targeting health-related behaviors (eg, diet, physical activity, and smoking and alcohol cessation).

The sources of the eHealth programs that were screened for eligibility are presented in Table 1. A total of 84 programs were randomly selected and rated between September and December 2016, following a systematic identification process, conducted as part of Enlight tool's development [21]. In this process, we created a list of mobile apps and Web-based programs through keyword searches in Google Play and Google search engine (this systematic process and a complete list of keywords is presented in Multimedia Appendix 1). Following the tool's development phase, we used a similar systematic identification process—that now also included paid programs—to rate additional programs between January and December 2017. All

programs found in the top 10 Web or mobile app search results (and that had not been rated before) were rated, reaching a number of additional 50 eHealth programs. Finally, 8 eHealth programs were identified based on recommendations from eHealth researchers and product developers (eg, the Digital Behavioral Health Workgroup at Northwell Health). This selection process yielded a list of 142 programs, among which, 21 programs had both mobile and Web-based versions. As the behavioral data on program usage and our inclusion criteria in this study relate separately to websites and mobile apps, we eventually screened for eligibility 71 Web-based programs and 92 mobile apps.

### Inclusion and Exclusion Criteria

To be included in this analysis, interventions had to be (1) self-guided and (2) cost-free. Apps that were free to install, yet had a trial period (ie, paid with free trial) were also included; however, we examined their impact on results using a sensitivity analysis, as will be further described in the data analysis section. Exclusion criteria included (1) programs that were only trackers, as their quality ratings do not rely on the full list of product ratings covered in Enlight; (2) websites that were not focused mostly on the intervention itself (eg, websites with many blog articles that did not require user log-in)—as the data on the platform are provided in an aggregated way that does not distinguish between the website's different Web pages; (3) apps that did not have an Android version, as the behavioral dataset did not include information on iOS apps; and (4) programs without usage data in our behavioral dataset (due to, for example, small number of users).

### Measures

#### Enlight Quality Ratings

Enlight is a comprehensive suite of criteria-based measurements—completed by trained raters who review the eHealth program—that was developed by a multidisciplinary team through a rigorous process of content analysis, grouping, and classification of 476 identified criteria items [21,40]. The tool covers 6 different product quality domains: Usability, Visual Design, User Engagement, Content, Therapeutic Persuasiveness (ie, persuasive design and incorporation of behavior change techniques), and Therapeutic Alliance (see Multimedia Appendix 2 for a detailed description of the scales used and operational definitions of all items). Each quality domain score ranges from 1 to 5 and is based on averaging the criteria ratings produced by the raters on a Likert scale (ranging from 1 to 5). For example, Therapeutic Persuasiveness is calculated by averaging the raters' scores of the following criteria items: call to action (eg, goal settings, prompting goals, and encouragement to complete goals), load reduction of activities (eg, set graded tasks), therapeutic rationale and pathway (eg, reasoned action and provide information about behavioral health link), rewards (eg, contingent rewards), real data-driven/adaptive content (eg, self-monitoring of behavior), ongoing feedback, and expectations and relevance.

**Table 1.** Sources of previously rated eHealth programs that were screened for eligibility in this study.

Source	eHealth programs (N=142), n (%)	Web-based program (N=71), n (%)	Mobile app (N=92), n (%)	Programs with 2 delivery mediums <sup>a</sup> (N=21), n (%)
Systematic review: programs randomly selected and rated during Enlight's development phase	84 (59.2)	43 <sup>b</sup> (61)	49 <sup>b</sup> (53)	8 (38)
Additional programs found in the top 10 mobile or Web search results of the systematic identification process	50 (35.2)	25 (35)	38 (41)	13 (62)
Additional programs that were identified based on personal recommendations	8 (5.6)	3 (4)	5 (5.4)	0 (0)

<sup>a</sup>The 2 delivery mediums are (1) Web-based programs and (2) mobile apps.

<sup>b</sup>In the original study, we examined 42 mobile apps and 42 Web-based programs. Eventually, 7 websites and 1 mobile app had a similar version in the other delivery medium (mobile or website).

Enlight covers 2 additional measures that were included in this analysis. Credibility consists of a checklist calculated by aggregating the scores received in each of its respective categorical items (owners' credibility, maintenance, strong advisory support, third party endorsement, and evidence of successful implementation). The checklist differs from the product quality assessments in 2 ways: (1) most credibility items cannot be rated before product deployment (as they rely on data that are collected afterward) and (2) the criteria included in the checklist are not expected to directly impact the end users' experience of the product's efficacy (however, they may expose the user to acknowledged risks or benefits). We also included an evidence-based program scale (ie, research evidence) that assesses the quality of the empirical research supporting the program within the current zeitgeist on a scale from 1 (very poor) to 5 (very good) [21].

A total of 3 individuals with clinical experience (eg, clinical psychologists) and with at least 1 year experience working in the eHealth domain (eg, content writing of eHealth programs and health technology coaching) performed the ratings, with 2 of them independently rating each program. Their training included a review of Enlight items, individual ratings of 7 eHealth programs that included group-solicited feedback on ratings, and then testing the ratings based on 5 additional programs (for a detailed review, see the study by Baumel et al [21]). In this study (post-training), the Enlight quality scales exhibited excellent inter-rater reliability scores between the 2 raters (intraclass correlations=.74 to .98, median .86).

### **Behavioral Data on User Engagement in the Real World**

Information on user behaviors was obtained from SimilarWeb's Pro panel data [41]. The panel provides aggregated nonpersonal information on user engagement with websites and mobile apps all over the world to enable Web and mobile app traffic research and analytics. It is based on several sources of anonymized usage data, such as that obtained from consenting users of mobile apps or browser add-ons (ie, products). A dedicated product team at SimilarWeb is responsible for building and partnering with hundreds of high-value consumer products that make up the panel. According to SimilarWeb, the products are used across diverse audiences, without cluttering the user with advertisements. Although benefiting from the products, users contribute to the panel, as they enable to document their online or mobile apps' usage activities seamlessly and anonymously

[41]. The data are not used by SimilarWeb or provided to any third parties for the purposes of marketing, advertising, or targeting any individual subjects. The data-gathering procedures comply with data privacy laws, including the way the data are collected, anonymized, stored, secured, and used. These procedures are updated regularly based on evolving data privacy legislation and requirements, such as the European Union's General Data Protection Regulation [42].

The study was approved by University of Haifa Institutional Review Board. The measures were set to include data gathered over an 18-month period from November 1, 2016 to April 30, 2018. To examine user engagement with mobile apps, we used 2 measures retrieved from SimilarWeb Pro: (1) average app usage time (of all users) and (2) the percentage of users who downloaded the app and were still using it 30 days later. As this information is provided separately per country, we examined the country with the most app downloads. User engagement with the Web-based interventions was calculated based on average monthly visits to the website multiplied by the average visit duration.

To test the reliability of the data for the remaining websites, we calculated the Spearman correlation between Web usage time obtained from the SimilarWeb platform and the usage time calculated for websites in our previous work with Microsoft Research [37]. A strong positive correlation ( $r=.69$ ;  $P<.001$ ) was found between the 2 datasets for the 17 websites that had data on both platforms. In light of the difference between the 2 datasets (our previous work was focused only on Explorer browser users), we also examined the Spearman correlation between website global popularity ranks (ie, a rank that reflects the extent to which a website is utilized all over the world) on SimilarWeb platform and Alexa [43]—an independent source of information on user traffic. A very strong Spearman correlation was found ( $n=28$ ;  $r=.78$ ;  $P<.001$ ). Relating to the validity of mobile app usage data, an Oath researcher [44] (RW) examined 30 randomly selected mobile apps with data on SimilarWeb and Oath's own independent records of mobile app usage data. The researcher examined the correlation between the average number of user sessions per day in the 2 datasets, finding a very strong Spearman correlation ( $n=30$ ;  $r=.77$ ;  $P<.001$ ). These findings suggested that there was sufficient convergent validity, which can be claimed if the correlation

coefficient is above .50, although a value above .70 is usually recommended [45].

Finally, we documented publicly available data that relate to a program's acceptability. The reported number of installs, average star ratings, and number of reviews were obtained from the Google Play store. Websites' total monthly visits were obtained from the publicly available version of SimilarWeb.

### Data Analysis

Medians and interquartile ranges (IQRs) were used to present the distribution of the variables. Pearson correlations were used to examine the relationships between the continuous variables. As Pearson correlations and linear regressions assume a normal distribution of the variables, we examined the skewness and kurtosis of the variables within acceptable limits of  $\pm 2$  and performed a logarithmic transformation (base 10) of these variables [46,47]. This transformation was eventually applied to the following variables: Credibility, Research Evidence, number of reviews on Google Play, mobile app usage time, and website's total monthly visits.

A sensitivity analysis was conducted to examine whether the pattern of correlations found differed depending on (1) the clinical aim of the intervention (ie, mental health or health-related behaviors) and (2) whether paid programs with a free trial were included or excluded. The differences between correlations of these groups were calculated using Fisher Z-transformation. Independent sample *t* tests with Bonferroni correction were performed to examine the difference between categorical items on the behavioral measures.

Finally, a series of hierarchical stepwise linear regressions was applied to examine the ability to predict user engagement in the real world independent of empirical testing. In the first step, we examined the percentage of variance explained by Enlight quality ratings. In the second step, we examined whether empirical data on real-world usage that were available at no cost (eg, Google Play for mobile apps: number of downloads, star reviews, and number of reviews; SimilarWeb for websites: monthly visits) significantly increased the explained variance. In each of these steps, a stepwise approach was applied to avoid adding predicting variables that did not contribute significantly to the overall model.

## Results

The analysis included user behavior data involving 52 mobile apps and 32 websites (see Figure 1 for the flow diagram of program selection and Multimedia Appendix 3 for a full list of included programs). Overall, 9 programs had usage data for both Web- and mobile-app versions, reaching a total of 75 different programs with data. Within the 18-month time window of the analysis, the 52 mobile apps had a median of 38,600 downloads (IQR 116,000), and the 32 Web-based interventions had a median of 5689 monthly unique visitors (IQR 30,038). The median monthly usage time was 5.11 min (IQR 20.51) for mobile apps and 9.0 min (IQR 7.37) for websites. The medians and IQRs of Enlight product quality ratings were as follows: Usability median 3.66 (IQR 1.66); Visual Design median 3.00

(IQR 1.00); User Engagement median 3.20 (IQR 1.10); Content median 3.50 (IQR 1.00); Therapeutic Persuasiveness median 2.71 (IQR 1.00); and Therapeutic Alliance median 2.66 (IQR 1.33).

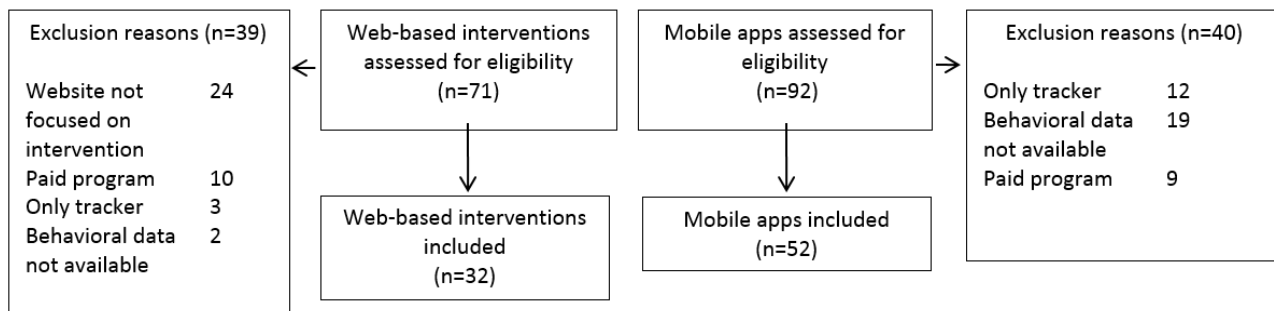
Pearson correlations between Enlight product quality ratings, Credibility, and Research Evidence are presented in Table 2. User Engagement, Content, Therapeutic Persuasiveness, and Therapeutic Alliance were positively correlated with Credibility ( $.26 \leq r \leq .51$ ,  $P \leq .01$ ) and Research Evidence ( $.21 \leq r \leq .39$ ,  $P \leq .04$ ). That is, the interventions with higher scores in these quality domains had higher Credibility and Research Evidence ratings.

Pearson correlations between Enlight quality ratings and items, Credibility and Research Evidence, publicly available data on program acceptance, and the 3 behavioral variables are presented in Table 3. Results point to moderate positive correlations between Therapeutic Persuasiveness, Therapeutic Alliance, and the 3 behavioral variables ( $.31 \leq r \leq .51$ ,  $P \leq .03$ ). Visual Design, User Engagement, and Content had a similar degree of correlations with the user engagement variables of mobile apps ( $.25 \leq r \leq .49$ ,  $P \leq .04$ ), but not Web-based interventions.

Altogether, a similar pattern of results was found between the criteria items of domains with significant correlations and the respective behavioral variables. It is worth noting that 3 criteria items of the Therapeutic Persuasiveness domain—rewards, data-driven/adaptive, and ongoing feedback—showed significant correlations with mobile app user retention after 30 days; however, these items did not correlate significantly with mobile app usage time. In terms of the Enlight checklists, a program's Credibility had positive correlations with the usage time of apps ( $r = .34$ ;  $P = .006$ ) and websites ( $r = .30$ ;  $P = .04$ ), but no significant correlations were found between Research Evidence and the 3 behavioral variables. Finally, results point to positive correlations between the number of reviews on Google Play and average app usage time ( $r = .58$ ;  $P < .001$ ) and user retention after 30 days ( $r = .23$ ;  $P = .049$ ). The number of installs and average star reviews had significant positive correlations with mobile app usage time ( $r = .25$  and  $.36$ , respectively;  $P = .04$  and  $.005$ , respectively).

As the Credibility checklist covers different independent categorized items, a series of independent sample *t* tests with Bonferroni correction was performed to examine the difference between Credibility items (Owners' Credibility, Maintenance, Strong Advisory Support, Evidence of Successful Implementation) in terms of the usage time of mobile and Web-based interventions. (Third Party Endorsement was not included in this test as a relevant party had endorsed only a few programs.) A significant difference was found in the average usage time of mobile apps, favoring programs that had Evidence of Successful Implementation (ie, a high number of downloads or positive reviews)  $t_{50} = 3.88$ ,  $P < .001$ , Cohen  $d = 1.07$ , and in maintenance, favoring programs that had been updated within the previous 6 months,  $t_{50} = 2.63$ ,  $P = .28$ , Cohen  $d = 0.73$ . As both variables relied on the data documented on Google Play, we added them in the second step of the regression analysis described below.

**Figure 1.** Flow diagram of program selection.



**Table 2.** Pearson correlations between Enlight product quality ratings, Credibility, and Research Evidence metrics (n=75).

Scale	Credibility		Research evidence	
	<i>r</i> <sup>a</sup>	<i>P</i> value	<i>r</i>	<i>P</i> value
Usability	.02	.45	.00	.50
Visual Design	.08	.24	-.13	.14
User Engagement	.39 <sup>b</sup>	<.001	.21 <sup>b</sup>	.03
Content	.51 <sup>b</sup>	<.001	.39 <sup>b</sup>	<.001
Therapeutic Persuasiveness	.26 <sup>b</sup>	.01	.21 <sup>b</sup>	.04
Therapeutic Alliance	.39 <sup>b</sup>	<.001	.26 <sup>b</sup>	.01

<sup>a</sup>Using Fisher Z-transformation, no significant differences in Pearson correlation values were found between programs targeting mental health (n=51) and those targeting health-related behaviors (n=24).

<sup>b</sup>Indicates significant correlations.

**Table 3.** Pearson correlations between Enlight scales, publicly available data on program acceptance, and real-world user engagement with eHealth interventions.

Scale	Mobile app interventions (n=52)				Web interventions (n=32)	
	Usage time		User 30-day retention		Usage time	
	<i>r</i> <sup>a</sup>	<i>P</i> value	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value
<b>Usability</b>	.15	.14	-.07	.32	.01	.48
Ease of use	.21	.07	.05	.36	.10	.29
Learnability	.07	.31	-.13	.18	-.14	.22
Navigation	.13	.18	-.11	.22	.03	.43
<b>Visual Design</b>	.31 <sup>b</sup>	.01	.25 <sup>b</sup>	.04	.09	.32
Aesthetics	.28 <sup>b</sup>	.02	.28 <sup>b</sup>	.02	.12	.26
Layout	.32 <sup>b</sup>	.01	.19	.09	.14	.22
Size	.24 <sup>b</sup>	.047	.18	.10	-.00	.49
<b>User Engagement</b>	.49 <sup>b</sup>	<.001	.39 <sup>b</sup>	.002	.21	.12
Captivating	.51 <sup>b</sup>	<.001	.31 <sup>b</sup>	.01	.38 <sup>b</sup>	.02
Content presentation	.53 <sup>b</sup>	<.001	.29 <sup>b</sup>	.02	.22	.12
Interactive	.13	.19	.25 <sup>b</sup>	.04	.28	.06
Not irritating	.23	.07	.17	.15	-.11	.31
Targeted/tailored/personalized	.41 <sup>b</sup>	.001	.38 <sup>b</sup>	.003	.01	.48
<b>Content</b>	.45 <sup>b</sup>	<.001	.33 <sup>b</sup>	.009	.16	.19
Evidence-based content	.43 <sup>b</sup>	.001	.31 <sup>b</sup>	.01	.15	.20
Information provision quality	.35 <sup>b</sup>	.006	.34 <sup>b</sup>	.006	-.01	.48
Complete and concise	.39 <sup>b</sup>	.002	.35 <sup>b</sup>	.005	.27	.07
Clarity about program's purpose	.43 <sup>b</sup>	.001	.10	.23	.09	.30
<b>Therapeutic Persuasiveness</b>	.36 <sup>b</sup>	.004	.39 <sup>b</sup>	.002	.33 <sup>b</sup>	.03
Call to action	.34 <sup>b</sup>	.007	.49 <sup>b</sup>	<.001	.29	.055
Load reduction of activities	.44 <sup>b</sup>	.001	.37 <sup>b</sup>	.004	.37 <sup>b</sup>	.02
Therapeutic rationale/pathway	.47 <sup>b</sup>	<.001	.36 <sup>b</sup>	.004	.21	.12
Rewards	.14	.16	.26 <sup>b</sup>	.03	.22	.12
Data-driven/adaptive	.11	.21	.24 <sup>b</sup>	.046	.23	.10
Ongoing feedback	.14	.16	.33 <sup>b</sup>	.01	.14	.23
Expectations and relevance	.31 <sup>b</sup>	.01	.03	.41	.19	.15
<b>Therapeutic Alliance</b>	.51 <sup>b</sup>	<.001	.31 <sup>b</sup>	.01	.36 <sup>b</sup>	.02
Basic acceptance and support	.45 <sup>b</sup>	<.001	.40 <sup>b</sup>	.002	.42 <sup>b</sup>	.009
Positive therapeutic expectations	.45 <sup>b</sup>	<.001	.33 <sup>b</sup>	.009	.22	.12
Relatability	.40 <sup>b</sup>	.002	.09	.26	.32 <sup>b</sup>	.04
<b>Enlight postempirical measures</b>						
Credibility	.34 <sup>b</sup>	.006	-.03	.42	.31 <sup>b</sup>	.04
Research evidence	.08	.28	-.13	.18	.19	.15

Scale	Mobile app interventions (n=52)				Web interventions (n=32)	
	Usage time		User 30-day retention		Usage time	
	<i>r</i> <sup>a</sup>	<i>P</i> value	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value
<b>Google Store data</b>						
Number of installs	.25 <sup>b</sup>	.04	.10	.25	N/A <sup>c</sup>	N/A
Average star reviews	.36 <sup>b</sup>	.005	.18	.11	N/A	N/A
Number of reviews	.58 <sup>b</sup>	<.001	.23 <sup>b</sup>	.049	N/A	N/A
<b>Web traffic data</b>						
Total monthly visits	N/A	N/A	N/A	N/A	.19	.15

<sup>a</sup>Using Fisher Z-transformation, no significant differences in Pearson correlation values were found between programs targeting mental health (mobile apps n=38 and websites n=22) and those targeting health-related behaviors (mobile apps n=14 and websites n=10) or between the full samples and the samples after subtracting paid programs with free trials (n=7).

<sup>b</sup>Indicates significant correlations.

<sup>c</sup>N/A: not applicable.

**Table 4.** Hierarchical stepwise regressions for predictors of user engagement with self-guided mobile and Web-based interventions.

Variable	B <sup>a</sup>	SE B	Beta	<i>P</i> value	R <sup>2</sup> change	<i>P</i> value
<b>Mobile app usage time (n=52)</b>						
<b>Step 1</b>						
Therapeutic Alliance	.47	.11	.51	<.001	.26	<.001
<b>Step 2</b>						
Therapeutic Alliance	.29	.11	.31	.02	.15	.001
Number of reviews	.29	.08	.44	.001	N/A <sup>b</sup>	N/A
<b>Mobile app 30-day retention (n=52)</b>						
<b>Step 1</b>						
Therapeutic Persuasiveness	.01	.01	.39	.004	.15	.004
Step 2 <sup>c</sup>	N/A	N/A	N/A	N/A	N/A	N/A
<b>Web intervention total usage time (n=32)</b>						
<b>Step 1</b>						
Therapeutic Alliance	2.87	1.34	.36	.04	.13	.04
Step 2 <sup>c</sup>	N/A	N/A	N/A	N/A	N/A	N/A

<sup>a</sup>B: unstandardized regression coefficient.

<sup>b</sup>N/A: not applicable.

<sup>c</sup>No new variables were entered into the equation.

Table 4 presents the hierarchical stepwise regressions performed to examine the predictors and percentages of real-world user engagement variance that were explained by Enlight product quality ratings and whether publicly available data on program acceptance added to the explained variance. The analysis showed that the variance of user engagement explained by Enlight ratings ranged between 13% and 26%. In each of the first steps, following the entrance of one Enlight quality rating—either Therapeutic Persuasiveness or Therapeutic Alliance—no other metric was found to significantly contribute to the model. The analysis also showed that publicly available data (number of reviews) explained 15% of the variance of mobile app usage

time above Enlight ratings; publicly available data did not significantly explain the variance of mobile app user retention or Web-based intervention usage time.

## Discussion

### Principal Findings

This study presents novel findings about the relationship between product design and user engagement with self-guided eHealth interventions in the real world. We found that product quality in terms of Visual Design, User Engagement, Content, Therapeutic Persuasiveness, and Therapeutic Alliance was



positively correlated with real-world usage of mobile apps. Therapeutic Persuasiveness and Therapeutic Alliance were also positively correlated with real-world usage of Web-based programs. Similar to previous findings, Visual Design, User Engagement, and Content were not found to have a significant correlation with usage time of Web-based interventions [37]. Although the domains' criteria items had similar patterns of correlation with the behavioral variables, 3 Therapeutic Persuasiveness items—rewards, data-driven/adaptive, and ongoing feedback—showed significant correlations with mobile app user retention after 30 days but no significant correlations with mobile app usage time. It might be that these variables, which examine the ongoing reciprocal interaction between the software and the user, are sensitive to the time window of use rather than to the accumulated usage time. However, future research should first be conducted to examine whether this pattern of results is replicated.

Congruent with our previous examination [37], Usability was not associated with the behavioral variables of user engagement. It is important to note that as our analysis was based on between-program evaluation, it does not mean that improving program's usability will not enhance user engagement with the program. Alternatively, we suggest that using the Usability score to compare *different* programs might not capture which programs are more engaging to users. Therefore, Usability should be considered to be a barrier rather than a facilitator of user engagement [48,49]. A future direction could be to define ranges—in terms of Usability scores—that identify the point at which a program is usable enough to be evaluated based on other metrics.

Our analysis of Google Play data revealed positive correlations between the number of reviews, mobile app usage time, and user retention; number of installs and average star reviews were also positively correlated with mobile app usage time. These findings present a link between data available to the public on app stores and an app's overall tendency to engage users. Regression results, however, suggested that these data do not always enhance our understanding of user engagement after accounting for the quality of product design—which could be determined before real-world use of the program. Hence, it could be informative to examine how expert-based rating tools such as Enlight can be used by developers during a program's development phase to guide the process of product design before empirical testing.

Our analysis also showed that programs with better design quality had better research evidence. At the same time, research evidence did not predict user engagement in real-world use. This finding is congruent with Fleming et al's study that examined published reports on the real-world uptake of eHealth programs [50]. The researchers found that indications of completion or sustained use (ie, completion of all modules or the last assessment or continuing to use apps after 6 weeks or more) varied from 0.5% to 28.6%, concluding that actual usage may vary from that reported in trials. Accordingly, our analysis implies that when research evidence supports a certain program, it does not necessarily mean that users will engage with the program in the real world. It is important to note that there may be many reasons for what could be referred to as “trial versus

real-world gap,” including the impact of the study setting on user engagement [25] and populations being targeted for the study that differ from those using the program in the real world [26]. Future research should be conducted to empirically examine this phenomenon and the factors influencing it.

It is also important to note that the analysis did not reveal an association between the Credibility items—developer's credibility (eg, academic institute) and strong advisory support group—and user engagement variables. One reason that could explain this finding is the high costs associated with app development [23], which may create barriers for teams that are more focused on answering academic research questions based on grant funding. However, this explanation should be further examined to draw firm conclusions.

Finally, from a methodological perspective, the notion that stands behind the development of criteria-based rating tools is that in the absence of proper training, proper expertise, and proper use of developed scales, inter-rater agreement will be low, and therefore, the examination will not be reliable. For example, in a study to examine inter-rater reliability of mobile health (mHealth) app rating measures, Powell et al gathered a panel of 6 reviewers to review the top 10 depression apps and top 10 smoking cessation apps from the Apple iTunes App Store and evaluate them based on several measures. The authors found a wide variation in the inter-rater reliability of measures, with some measures being more robust across categories of apps than others [36]. In recent years, several studies have demonstrated that it is feasible to achieve sufficient inter-rater reliability with criteria-based rating scales by using trained raters [35,51]. To perform the ratings for this study and previous evaluations using the Enlight tool, raters had to complete a certain level of training to provide reliable ratings. This notion has been acknowledged for decades within the psychological assessment field ([52,53]; George et al, unpublished data, 1985) we hope that as the eHealth evaluation field moves forward, more attention will be paid to the proper use of methods to train and examine evaluators' work.

## Limitations

This study has several limitations. First, the findings are not based on an experimental procedure that compared different designs of the *same* intervention. However, it would not be possible to utilize an experimental procedure to compare many aspects of product design at the same time. Consequently, if an association is not found between a quality domain and program usage, it does not mean that improving the program in this domain will not influence usage. Instead, it means that when comparing *different* programs, some aspects of quality are more important than others in predicting usage time. Second, this study examined user engagement, which is not the same as efficacy. Data suggest that there is a strong relationship between engagement and efficacy [54-56]; however, more does not always equal better [57]. A future research direction would be to measure efficacy using a large sample of programs “in the real world” and to examine the correlations between product design and efficacy. This testing should take into account fundamental questions related to participant consent and how to measure intervention outcomes.

## Conclusions

Results indicate that Enlight is a valid tool for predicting real-world user engagement with eHealth interventions, based on expert evaluations that were conducted before empirical testing with end users. The link between expert reviews and user behaviors in the real world supports the importance of rating tools that may enable trained experts to (1) guide the design of evidence-based interventions before testing with end users and (2) provide important details about the product's potential to enable end users to make educated decisions when

searching for self-guided interventions. Such details are presented in MindTools.io [39], a nonprofit website that publishes in-depth app reviews using Enlight rating scales.

Finally, the use of real-world behavioral datasets, which are garnered from a massive number of users, is a novel way to learn about user behaviors, creating new avenues of research and advancing our understanding of eHealth interventions. More studies are needed to shed light on the relationships between real-world uptake and data that emerge from other sources of information.

## Acknowledgments

AB and JMK's work on eHealth rating scales and the development of MindTools.io is supported by the Donald & Barbara Zucker Family Foundation. The authors would like to thank Dr Ran Wolff for his support in examining the validity of the behavioral user panel they were using.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Description of the systematic identification process.

[\[PDF File \(Adobe PDF File\), 60KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Description of scales and scale items.

[\[PDF File \(Adobe PDF File\), 144KB-Multimedia Appendix 2\]](#)

## Multimedia Appendix 3

Included programs.

[\[PDF File \(Adobe PDF File\), 62KB-Multimedia Appendix 3\]](#)

## References

1. Kazdin AE. Addressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. *Behav Res Ther* 2017 Dec;88:7-18. [doi: [10.1016/j.brat.2016.06.004](https://doi.org/10.1016/j.brat.2016.06.004)] [Medline: [28110678](https://pubmed.ncbi.nlm.nih.gov/28110678/)]
2. Baumel A, Baker J, Birnbaum ML, Christensen H, De Choudhury M, Mohr DC, et al. Summary of key issues raised in the Technology for Early Awareness of Addiction and Mental Illness (TEAAM-I) Meeting. *Psychiatr Serv* 2018 May 01;69(5):590-592. [doi: [10.1176/appi.ps.201700270](https://doi.org/10.1176/appi.ps.201700270)] [Medline: [29334875](https://pubmed.ncbi.nlm.nih.gov/29334875/)]
3. Eysenbach G. The law of attrition. *J Med Internet Res* 2005;7(1):e11 [FREE Full text] [doi: [10.2196/jmir.7.1.e11](https://doi.org/10.2196/jmir.7.1.e11)] [Medline: [15829473](https://pubmed.ncbi.nlm.nih.gov/15829473/)]
4. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. *J Med Internet Res* 2009;11(2):e13 [FREE Full text] [doi: [10.2196/jmir.1194](https://doi.org/10.2196/jmir.1194)] [Medline: [19403466](https://pubmed.ncbi.nlm.nih.gov/19403466/)]
5. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. *Telemed J E Health* 2009;15(3):231-240. [doi: [10.1089/tmj.2008.0099](https://doi.org/10.1089/tmj.2008.0099)] [Medline: [19382860](https://pubmed.ncbi.nlm.nih.gov/19382860/)]
6. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. *Am J Prev Med* 2007 Oct;33(4):336-345 [FREE Full text] [doi: [10.1016/j.amepre.2007.05.007](https://doi.org/10.1016/j.amepre.2007.05.007)] [Medline: [17888860](https://pubmed.ncbi.nlm.nih.gov/17888860/)]
7. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. *J Ment Health* 2015 Aug 19;24(5):321-332. [doi: [10.3109/09638237.2015.1019054](https://doi.org/10.3109/09638237.2015.1019054)] [Medline: [26017625](https://pubmed.ncbi.nlm.nih.gov/26017625/)]
8. Baumeister RF, Vohs KD. Self-Regulation, ego-depletion, and motivation. *Soc Personal Psychol Compass* 2007;1(1):115-128. [doi: [10.1111/j.1751-9004.2007.00001](https://doi.org/10.1111/j.1751-9004.2007.00001)]
9. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. *J Med Internet Res* 2013;15(7):e146 [FREE Full text] [doi: [10.2196/jmir.2665](https://doi.org/10.2196/jmir.2665)] [Medline: [23859884](https://pubmed.ncbi.nlm.nih.gov/23859884/)]

10. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. *Ann Behav Med* 2009 Aug;38(1):18-27 [FREE Full text] [doi: [10.1007/s12160-009-9133-4](https://doi.org/10.1007/s12160-009-9133-4)] [Medline: [19802647](https://pubmed.ncbi.nlm.nih.gov/19802647/)]
11. Kok G, Schaalma H, Ruiters RA, van Empelen P, Brug J. Intervention mapping: protocol for applying health psychology theory to prevention programmes. *J Health Psychol* 2004 Jan;9(1):85-98. [doi: [10.1177/1359105304038379](https://doi.org/10.1177/1359105304038379)] [Medline: [14683571](https://pubmed.ncbi.nlm.nih.gov/14683571/)]
12. Provost M, Koopalum D, Dong D, Martin BC. The initial development of the WebMedQual scale: domain assessment of the construct of quality of health web sites. *Int J Med Inform* 2006 Jan;75(1):42-57. [doi: [10.1016/j.ijmedinf.2005.07.034](https://doi.org/10.1016/j.ijmedinf.2005.07.034)] [Medline: [16169770](https://pubmed.ncbi.nlm.nih.gov/16169770/)]
13. Muench F, Baumel A. More than a text message: dismantling digital triggers to curate behavior change in patient-centered health interventions. *J Med Internet Res* 2017 Dec 26;19(5):e147 [FREE Full text] [doi: [10.2196/jmir.7463](https://doi.org/10.2196/jmir.7463)] [Medline: [28550001](https://pubmed.ncbi.nlm.nih.gov/28550001/)]
14. 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(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/)]
15. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. *J Med Internet Res* 2012;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/)]
16. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *J Med Internet Res* 2010;12(1):e4 [FREE Full text] [doi: [10.2196/jmir.1376](https://doi.org/10.2196/jmir.1376)] [Medline: [20164043](https://pubmed.ncbi.nlm.nih.gov/20164043/)]
17. Bickmore T, Schulman D, Yin L. Maintaining engagement in long-term interventions with relational agents. *Appl Artif Intell* 2010 Jul 01;24(6):648-666 [FREE Full text] [doi: [10.1080/08839514.2010.492259](https://doi.org/10.1080/08839514.2010.492259)] [Medline: [21318052](https://pubmed.ncbi.nlm.nih.gov/21318052/)]
18. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. *Internet Interv* 2017;10:39-46. [doi: [10.1016/j.invent.2017.10.002](https://doi.org/10.1016/j.invent.2017.10.002)] [Medline: [30135751](https://pubmed.ncbi.nlm.nih.gov/30135751/)]
19. Cavanagh K, Millings A. (Inter) personal computing: the role of the therapeutic relationship in e-mental health. *J Contemp Psychother* 2013;43(4):197-206. [doi: [10.1007/s10879-013-9242-z](https://doi.org/10.1007/s10879-013-9242-z)]
20. Holter MT, Johansen A, Brendryen H. How a fully automated eHealth program simulates three therapeutic processes: a case study. *J Med Internet Res* 2016 Jun 28;18(6):e176 [FREE Full text] [doi: [10.2196/jmir.5415](https://doi.org/10.2196/jmir.5415)] [Medline: [27354373](https://pubmed.ncbi.nlm.nih.gov/27354373/)]
21. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. *J Med Internet Res* 2017 Mar 21;19(3):e82 [FREE Full text] [doi: [10.2196/jmir.7270](https://doi.org/10.2196/jmir.7270)] [Medline: [28325712](https://pubmed.ncbi.nlm.nih.gov/28325712/)]
22. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *Am J Prev Med* 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: [10.1016/j.amepre.2007.01.022](https://doi.org/10.1016/j.amepre.2007.01.022)] [Medline: [17466815](https://pubmed.ncbi.nlm.nih.gov/17466815/)]
23. Research2Guidance. It costs \$425,000 on average to develop an mHealth app – Download Free Whitepaper URL: <https://research2guidance.com/it-costs-425000-dollars-on-average-to-develop-an-mhealth-app-download-free-whitepaper/> [accessed 2018-06-24] [WebCite Cache ID 70PgZ4yuP]
24. Moller AC, Merchant G, Conroy DE, West R, Hekler E, Kugler CK, et al. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. *J Behav Med* 2016;40(1):85-98. [doi: [10.1007/s10865-016-9818-7](https://doi.org/10.1007/s10865-016-9818-7)]
25. Ebert DD, Baumeister H. Internet-based self-help interventions for depression in routine care. *JAMA Psychiatry* 2017 Aug 01;74(8):852-853. [doi: [10.1001/jamapsychiatry.2017.1394](https://doi.org/10.1001/jamapsychiatry.2017.1394)] [Medline: [28636717](https://pubmed.ncbi.nlm.nih.gov/28636717/)]
26. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research . . . and three things we can do about them. *Psychiatr Serv* 2017 May 01;68(5):427-429. [doi: [10.1176/appi.ps.201600541](https://doi.org/10.1176/appi.ps.201600541)] [Medline: [28412890](https://pubmed.ncbi.nlm.nih.gov/28412890/)]
27. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smartphone health apps. *Telemed J E Health* 2015 Feb;21(2):97-104. [doi: [10.1089/tmj.2014.0088](https://doi.org/10.1089/tmj.2014.0088)] [Medline: [25469795](https://pubmed.ncbi.nlm.nih.gov/25469795/)]
28. Madan A, Dubey S. Usability evaluation methods: a literature review. *Int J Eng Sci* 2012;4(2):590-599 [FREE Full text]
29. Kientz JA, Choe EK, Birch B, Maharaj R, Fonville A, Glasson C, et al. Heuristic evaluation of persuasive health technologies. 2010 Presented at: Proceedings of the 1st ACM International Health Informatics Symposium; November 11-12, 2010; Arlington, VA, USA.
30. Monkman H, Griffith J, Kushniruk AW. Evidence-based heuristics for evaluating demands on eHealth literacy and usability in a mobile consumer health application. *Stud Health Technol Inform* 2015;216:358-362. [doi: [10.3233/978-1-61499-564-7-358](https://doi.org/10.3233/978-1-61499-564-7-358)] [Medline: [26262071](https://pubmed.ncbi.nlm.nih.gov/26262071/)]
31. Nielsen J. Enhancing the explanatory power of usability heuristics. 1994 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 24 - 28, 1994; Boston, MA, USA.
32. Karat CM. A Comparison of user interface evaluation methods. In: Nielsen J, Mack RL, editors. Usability inspection methods. New York: John Wiley & Sons; 1994:203-233.

33. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* 1999 Feb;53(2):105-111 [[FREE Full text](#)] [Medline: [10396471](#)]
34. Olsina L, Rossi G. Measuring web application quality with WebQEM. *IEEE Multimed* 2002;9(4):20-29. [doi: [10.1109/MMUL.2002.1041945](#)]
35. 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 Mar 11;3(1):e27 [[FREE Full text](#)] [doi: [10.2196/mhealth.3422](#)] [Medline: [25760773](#)]
36. Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, et al. Interrater reliability of mHealth app rating measures: analysis of top depression and smoking cessation apps. *JMIR Mhealth Uhealth* 2016;4(1):e15 [[FREE Full text](#)] [doi: [10.2196/mhealth.5176](#)] [Medline: [26863986](#)]
37. Baumel A, Yom-Tov E. Predicting user adherence to behavioral eHealth interventions in the real world: examining which aspects of intervention design matter most. *Transl Behav Med* 2018 Sep 08;8(5):793-798. [doi: [10.1093/tbm/ibx037](#)] [Medline: [29471424](#)]
38. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. *J Med Internet Res* 2017 Apr 19;19(4):e125 [[FREE Full text](#)] [doi: [10.2196/jmir.5604](#)] [Medline: [28428170](#)]
39. MindTools.io. URL: <https://mindtools.io/> [accessed 2018-06-24] [[WebCite Cache ID 70Pg0KTXJ](#)]
40. Baumel A, Birnbaum ML, Sucala M. A systematic review and taxonomy of published quality criteria related to the evaluation of user-facing eHealth programs. *J Med Syst* 2017 Aug;41(8):128. [doi: [10.1007/s10916-017-0776-6](#)] [Medline: [28735372](#)]
41. SimilarWeb. URL: <https://www.similarweb.com> [accessed 2018-06-18] [[WebCite Cache ID 70GZ6eAt5](#)]
42. European Commission. Data protection: Rules for the protection of personal data inside and outside the EU URL: [https://ec.europa.eu/info/law/law-topic/data-protection\\_en](https://ec.europa.eu/info/law/law-topic/data-protection_en) [accessed 2018-06-18] [[WebCite Cache ID 70GY1yLYp](#)]
43. Alexa. URL: <http://www.alexa.com/> [accessed 2018-01-04] [[WebCite Cache ID 6nGSLq16u](#)]
44. Oath. URL: <https://www.oath.com/> [accessed 2018-10-14] [[WebCite Cache ID 73A1knb5w](#)]
45. Carlson KD, Herdman AO. Understanding the impact of convergent validity on research results. *Organ Res Methods* 2012;15(1):17-32. [doi: [10.1177/1094428110392383](#)]
46. Gravetter F, Wallnau L. *Statistics for the Behavioral Sciences*. Belmont, CA, USA: Cengage Learning; 2012.
47. George D, Mallery P. *IBM SPSS Statistics 23 step by step: A simple guide and reference*. New-York, NY, USA: Routledge; 2016.
48. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. *J Assoc Inf Sci Technol* 2008;59(6):938-955. [doi: [10.1002/asi.20801](#)]
49. Atkinson M, Kydd C. Individual characteristics associated with World Wide Web use: an empirical study of playfulness and motivation. *ACM SIGMIS Database* 1997;28(2):53-62. [doi: [10.1145/264701.264705](#)]
50. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. *J Med Internet Res* 2018 Jun 06;20(6):e199 [[FREE Full text](#)] [doi: [10.2196/jmir.9275](#)] [Medline: [29875089](#)]
51. Dobbins M, Watson S, Read K, Graham K, Nooraie RY, Levinson AJ. A tool that assesses the evidence, transparency, and usability of online health information: development and reliability assessment. *JMIR Aging* 2018;1(1):e3. [doi: [10.2196/aging.9216](#)]
52. Lobbestael J, Leurgans M, Arntz A. Inter-rater reliability of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I) and Axis II Disorders (SCID II). *Clin Psychol Psychother* 2011;18(1):75-79. [doi: [10.1002/cpp.693](#)] [Medline: [20309842](#)]
53. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 1997 Jul;36(7):980-988. [doi: [10.1097/00004583-199707000-00021](#)] [Medline: [9204677](#)]
54. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. *J Med Internet Res* 2011;13(1):e17. [doi: [10.2196/jmir.1367](#)] [Medline: [21320854](#)]
55. Danaher BG, Smolkowski K, Seeley JR, Severson HH. Mediators of a successful web-based smokeless tobacco cessation program. *Addiction* 2008 Oct;103(10):1706-1712 [[FREE Full text](#)] [doi: [10.1111/j.1360-0443.2008.02295.x](#)] [Medline: [18715238](#)]
56. Makai P, Perry M, Robben SH, Schers HJ, Heinen MM, Olde Rikkert MG, et al. Evaluation of an eHealth intervention in chronic care for frail older people: why adherence is the first target. *J Med Internet Res* 2014;16(6):e156 [[FREE Full text](#)] [doi: [10.2196/jmir.3057](#)] [Medline: [24966146](#)]
57. Christensen H, Griffiths K, Groves C, Korten A. Free range users and one hit wonders: community users of an internet-based cognitive behaviour therapy program. *Aust N Z J Psychiatry* 2006 Jan;40(1):59-62. [doi: [10.1080/j.1440-1614.2006.01743.x](#)] [Medline: [16403040](#)]

---

**Abbreviations****eHealth:** electronic health**IQR:** interquartile range**mHealth:** mobile health

---

*Edited by G Eysenbach; submitted 05.07.18; peer-reviewed by S Schueller; comments to author 05.08.18; revised version received 16.09.18; accepted 22.10.18; published 14.12.18*

*Please cite as:*

*Baumel A, Kane JM*

*Examining Predictors of Real-World User Engagement with Self-Guided eHealth Interventions: Analysis of Mobile Apps and Websites Using a Novel Dataset*

*J Med Internet Res 2018;20(12):e11491*

*URL: <http://www.jmir.org/2018/12/e11491/>*

*doi: [10.2196/11491](https://doi.org/10.2196/11491)*

*PMID: [30552077](https://pubmed.ncbi.nlm.nih.gov/30552077/)*

©Amit Baumel, John M Kane. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 14.12.2018. 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.