Published in final edited form as:

Assessment. 2022 September ; 29(6): 1331–1345. doi:10.1177/10731911211015310.

Alliance with an Unguided Smartphone App: Validation of the Digital Working Alliance Inventory

Simon B. Goldberg

University of Wisconsin – Madison

Scott A. Baldwin

Brigham Young University

Kevin M. Riordan

University of Wisconsin – Madison

John Torous

Harvard University

Cortland J. Dahl,

Richard J. Davidson,

Matthew J. Hirshberg

University of Wisconsin – Madison

Abstract

The working alliance may be relevant in unguided smartphone-based interventions, but no validated measure exists. We evaluated the psychometric properties of the six-item Digital Working Alliance Inventory (DWAI) using a cross-sectional survey of meditation app users (n = 290) and the intervention arm of a randomized trial testing a smartphone-based meditation app (n = 314). Exploratory factor analysis suggested a single factor solution which was replicated using longitudinal confirmatory factor analysis. The DWAI showed adequate internal consistency and test-retest reliability. Discriminant validity was supported by a lack of association with social desirability, psychological distress, and preference for a waitlist condition. Convergent validity was supported by positive associations with perceived app effectiveness and preference for an app condition. Supporting predictive validity, DWAI scores positively predicted self-reported and objective app utilization. When assessed at weeks 3 or 4 of the intervention, but not earlier, DWAI scores predicted pre-post reductions in psychological distress.

Correspondence should be addressed to: Simon B. Goldberg, Department of Counseling Psychology, University of Wisconsin-Madison, 335 Education Building, 1000 Bascom Mall, Madison, Wisconsin, 53706, United States, phone: 608-265-8986, fax: 608-265-4174, sbgoldberg@wisc.edu.

Simon B. Goldberg, Department of Counseling Psychology and Center for Healthy Minds, University of Wisconsin-Madison, Madison, WI, USA; Scott A. Baldwin, Department of Psychology, Brigham Young University, Provo, UT, USA; Kevin M. Riordan, Department of Counseling Psychology and Center for Healthy Minds, University of Wisconsin-Madison, Madison, WI, USA; John Torous, Division of Digital Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA; Cortland J. Dahl, Center for Healthy Minds, University of Wisconsin-Madison and Healthy Minds Innovations, Inc., Madison, WI, USA; Richard J. Davidson, Center for Healthy Minds, Department of Psychology, and Department of Psychiatry, University of Wisconsin-Madison, Madison, WI, USA; Matthew J. Hirshberg, Center for Healthy Minds, University of Wisconsin-Madison, Madison, WI, USA.

Keywords

working alliance; digital technology; smartphone-based interventions; mobile health; validation study

Based on over four decades of research, the therapeutic or working alliance between patient and therapist has emerged as one of the most widely studied and robust predictors of treatment outcome in psychotherapy (Flückiger et al., 2018; Horvath & Symonds, 1991). Most modern measures of the alliance in psychotherapy derive from Bordin's (1979) conceptualization which includes agreement between patient and therapist on the tasks and goals of treatment along with an emotional bond characterized by trust and acceptance (e.g., Horvath & Greenberg, 1989). The largest meta-analysis of the alliance in adult psychotherapy, based on 295 studies and over 30,000 patients, detected an association between alliance and outcome of $r = .278$ (Flückiger et al., 2018). Notably, the alliance-outcome association did not differ across treatment types (e.g., cognitive behavioral therapy vs. psychodynamic therapy), specific alliance measure, or rater perspective (e.g., patient vs. therapist). Though the causal influence of alliance on outcome has not been definitively established due to methodological and ethical barriers (e.g., inability to randomly assign patients to an intentionally low-alliance psychotherapy), the alliance-outcome association appears relatively independent of patients' intake characteristics (e.g., symptom severity) and other therapeutic processes which may account for the apparent alliance-outcome association (e.g., therapist competence; Flückiger et al., 2020).

A wide variety of measures have been developed to assess this construct in psychotherapy. Although Flückiger et al.'s (2018) meta-analysis included 39 distinct measures of alliance, the majority (69%) were based on the Working Alliance Inventory (WAI; Horvath & Greenberg, 1989). The original WAI included 36 items and yielded three highly correlated subscales (rs ≥ .69; Horvath & Greenberg, 1989) corresponding to Bordin's (1979) task (e.g., "I believe the way we are working on my problem is correct"), goal (e.g., "I feel that the things I do in therapy will help me to accomplish the changes that I want"), and bond dimensions (e.g., "I believe ____ likes me"). Early factor analysis of the measure confirmed three distinct factors along with an overarching alliance factor (i.e., hierarchical factor model; Tracey & Kokotovic, 1989). A number of subsequent studies have evaluated factor structure for the WAI and its revisions which include several short forms, typically supporting either a three-factor solution or a two-factor solution with the Task and Goal subscales combined into a single factor (see Falkenström, Hatcher, & Holmqvist, 2015a). Recently, Falkenström et al. (2015b) developed a six-item version of the measure which they recommend be treated as unidimensional.

Alliance in Mobile Health

Face-to-face interventions upon which the bulk of the research on alliance is based have long been the gold standard delivery modality for psychotherapy. Following rapid technological advances in the past two decades, there is growing interest in the use of mobile health (mHealth) technology (e.g., the use of health-related smartphone apps) to expand access to mental health care (Aboujaoude et al., 2015) and to reduce mental health

inequity (Anderson-Lewis et al., 2018). Enthusiasm for mHealth interventions has only grown during the COVID-19 pandemic, as quarantine and social distancing measures taken to slow disease transmission have required movement away from traditional, in-person delivery of psychotherapy (Liu et al., 2020; Torous et al., 2020; Zhou et al., 2020). A key dimension that differentiates between mHealth interventions is the degree of therapist involvement, which can be high (e.g., traditional, synchronous 50-minute psychotherapy session delivered over phone or video; Osenbach et al., 2013), low (e.g., Internet-based cognitive behavioral therapy paired with coaching calls; Gilbody et al., 2015), or non-existent (e.g., unguided smartphone-based interventions; Weisel et al., 2020). Clearly, the amount of therapist involvement influences the cost and scalability of a given mHealth intervention. Although some degree of meta-analytic evidence exists supporting mHealth interventions which range in therapist involvement from traditional telehealth to unguided smartphone apps (Osenbach et al., 2013; Firth et al., 2017a, 2017b; Weisel et al., 2020), it does appear that therapist support increases efficacy (Linardon et al., 2019). Moreover, minimally guided and unguided mHealth interventions have notoriously high and rapid rates of disengagement (Baumel & Kane, 2018; Baumel et al., 2019; Eysenbach, 2005; Linardon & Fuller-Tyszkiewicz, 2020; Pratap et al., 2020).

Against the backdrop of promising efficacy, unmet clinical need, and rapid disengagement, researchers have begun applying theoretical frameworks drawn from the psychotherapy literature (e.g., object relations; Cohen & Torous, 2019) as potential solutions for increasing efficacy and engagement in mHealth interventions. Given its centrality for in-person psychotherapy, one may naturally ask what becomes of the alliance in mHealth interventions (Aboujaoude et al., 2015; Berger, 2017; Wehmann et al., 2020). Moreover, to the extent that alliance is relevant, it may help to explain low engagement in mHealth interventions and ultimately be harnessed to increase their efficacy.

In support of a digital corollary of the alliance, Flückiger et al.'s (2018) meta-analysis included 23 samples investigating the alliance-outcome in guided mHealth interventions (primarily Internet-based cognitive behavioral therapy), detecting an association almost identical to that found for in-person psychotherapy ($r = .275$). Researchers have recently also begun developing and testing measures of alliance for unguided mHealth modalities. Two studies (Miloff et al., 2020; Miragall et al., 2015) adapted the WAI for virtual reality and/or augmented reality therapies, primarily by replacing "my therapist" with "the virtual environment" or "the virtual therapist." Both studies showed the expected association between their adapted WAI and outcomes. Herrero et al. (2020) also adapted the WAI primarily by replacing "my therapist" with "the program" and used it to assess the online component of a blended face-to-face and Internet-based cognitive behavioral therapy, also showing the expected alliance-outcome association. Similarly, Kiluk et al. (2014) used an adapted version of the WAI to evaluate alliance for the computer-based component of a multicomponent intervention, although their measure did not predict outcomes (cocaine use).

These initial measure development efforts suggest some digital corollary of the alliance may exist, or at least something conceptually similar that also may predict treatment outcomes. While promising, the available measures may have important limitations. In particular,

measures adapting the WAI by replacing “my therapist” with reference to non-human technology can result in strange items (e.g., “The program and I respect each other”; Herrero et al., 2020). Both clinicians and patients have criticized anthropomorphic items (Berry et al., 2018). This issue may become especially salient when alliance is being assessed in a fully unguided context (e.g., unguided smartphone app). In addition, prior studies focused on alliance in unguided mHealth have involved some measure of human support (e.g., Kiluk et al., 2014) or interaction with a virtual therapist (e.g., Miloff et al., 2020). Prior studies on the alliance in unguided mHealth interventions have also been limited in sample size, prohibiting the use of both exploratory and confirmatory factor analysis and have not fully examined key psychometric characteristics (e.g., test-retest reliability, discriminant validity).

To our knowledge, no validated measure exists specifically designed to assess alliance within the context of a fully unguided mHealth intervention. Henson et al. (2019) proposed the Digital Working Alliance Inventory (DWAI) in their review of alliance in smartphone interventions for serious mental illness. Like prior efforts, Henson et al. took the WAI as their starting point, including items from the Task, Goals, and Bond subscales. But, informed by qualitative evaluations (Berry et al., 2018), items were adapted to avoid anthropomorphizing technology while retaining a sense of human connection with the app content. The measure was also kept to six items, in keeping with existing psychometrically sound short forms of the WAI (Falkenström et al., 2015b).

The Current Study

We sought to evaluate the psychometric properties of the DWAI for use as a measure of alliance within the context of unguided smartphone interventions. To do so, we used data drawn from two studies. Both studies were focused on smartphone-based meditation apps. Although this is only one of a wide variety of types of mental health apps that exist, meditation apps represent the vast majority of both daily and monthly active mental health app use (Wasil et al., 2020). Study 1 involved a cross-sectional online survey assessing the DWAI in reference to various smartphone-based meditation apps participants have used. Study 2 included data drawn from the intervention arm of a randomized controlled trial testing a specific smartphone-based meditation app. As no previous evaluation exists, we aimed to establish the measure’s structure through both exploratory and confirmatory factor analysis. We aimed to evaluate both internal consistency and test-retest reliability. We also planned to assess discriminant validity with unrelated constructs (e.g., social desirability, pre-treatment psychological symptoms, pre-treatment preference for a waitlist condition), convergent validity with related constructs (e.g., pre-treatment preference for the smartphone-based intervention, perceived app effectiveness), and predictive validity for changes in psychological distress. In order to investigate the possibility that alliance may be used to predict disengagement, we also examined associations between alliance and app utilization as an additional form of predictive validity.

Method

Participants and Procedure

Study 1 involved a cross-sectional online survey. Participants for Study 1 were recruited through Prolific (www.Prolific.co). Similar to functionality available through Amazon's Mechanical Turk, Prolific is an online participant recruitment platform. Documented advantages of Prolific include access to a pool of more diverse participants who are both less dishonest (i.e., less likely to cheat in order to gain a bonus payment) and more naïve (i.e., unfamiliar with commonly used measures) than Mechanical Turk (Peer et al., 2017). Participants in Study 1 were drawn from a larger study investigating the utilization of meditation in the United States population. Eligible participants for the larger study were adults (≥18 years old) living in the United States. The current analyses were not part of the preregistered aims of the larger study (https://osf.io/4h86s/?view_only=0e5d7ad85f87468ea40e047b3cf7c795). We include data from participants who passed two attention check items ("please select the leftmost response", "I have been randomly selecting responses on this survey"), reported having used a smartphone-based meditation app, and completed the DWAI in relation to their most used meditation app (n = 290). Consistent with the general population (Wasil et al., 2020), the two most commonly used apps were Headspace (31.0%) and Calm (28.6%). Both apps include instruction in various meditation techniques (e.g., focusing attention on the breath, scanning attention through the body, generating feelings of care for others) with content focused on increasing wellbeing and attention regulation as well decreasing stress, anxiety, and sleep disruption. Participants in Study 1 were on average 39.93 years old (SD = 14.77); 56.2% (n = 163) were female, 42.1% (n = 122) male, and 1.7% (n = 5) gender non-binary; 70.3% (n = 204) were non-Hispanic White, 12.4% (n = 36) Black, 7.9% (n = 23) Latinx, 6.2% (n = 18) Asian, 2.8% (n = 8) multiracial, and 0.3% (n = 1) Native American; 60.0% (n = 174) had completed college; 59.0% (n = 171) had an annual income ≤$50,000. The survey was administered between November 22nd and December 4th, 2020.

Participants for Study 2 were drawn from the intervention arm of a randomized controlled trial testing a smartphone-based meditation app – the Healthy Minds Program (HMP; for a description of the app and its underlying model of well-being, see Dahl et al., 2020; Goldberg et al., 2020b) in comparison to a waitlist control (https://osf.io/eqgt7). The HMP app includes four modules with practices aimed at developing skills supportive of wellbeing. These include practices to strengthen mindfulness and stabilize attention (Awareness), cultivate qualities like appreciation and kindness that contribute to healthy relationships with oneself and others (Connection), support self-inquiry (Insight), and enhance meaning in life by clarifying and applying core values and self-transcendent motivations (Purpose). The current analyses do not overlap with the primary aims of the larger trial, although we did hypothesize in separate preregistrations that the DWAI would be associated with changes in outcomes (https://osf.io/85kya/?view_only=cef24dc17c784a9790e388d5b5814f1d) and adherence (https://osf.io/swerk/?view_only=b31cd287334b491d806d0c63f3a583fb). Results from the full set of preregistered analyses may be published elsewhere. Eligible participants for the larger trial were adults (≥18 years old) currently employed by a school district within the state of Wisconsin. Additional exclusion criteria included extensive meditation

experience (i.e., retreat experience, weekly practice for ≥1 year, daily practice for ≥6 months), past use of the HMP app, and lack of access to a device capable of running the HMP app. Outcomes were assessed at baseline and post-treatment (4-weeks post-baseline). The DWAI was administered one, two, three, and four-weeks post-baseline for those in the HMP condition. The current study includes those participants randomized to the HMP condition who completed the DWAI at least once (n = 314). Participants in Study 2 were on average 42.66 years old (SD = 11.00); 84.7% (n = 266) were female, 11.5% (n = 36) male, and 0.3% (n = 1) gender non-binary, and 3.5% (n = 11) of unknown gender; 86.0% (n = 270) were non-Hispanic White, 2.6% (n = 8) Black, 0.3% (n = 1) Latinx, 1.3% (n = 4) Asian, 4.1% (n = 13) multiracial, 0.3% (n = 1) Native American, and 5.4% (n = 17) of unknown race/ethnicity; 86.0% (n = 270) had completed college; 15.3% (n = 48) had an annual income ≤$50,000. Recruitment for Study 2 occurred between June 18th and September 7th, 2020.

Measures

Alliance.—The DWAI (Henson et al., 2019) was used to assess digital working alliance in both Study 1 and Study 2. This six-item measure is based on items included in the WAI (Horvath & Greenberg, 1989), but adapted for the context of an unguided mHealth delivery format. Two items are included from each of the three WAI domains: Task (Item 2: "I believe the app tasks will help me to address my problem," Item 5: "The app is easy to use and operate"), Goal (Item 1: "I trust the app to guide me towards my personal goals," Item 4: "I agree that the tasks within the app are important for my goals"), and Bond (Item 3: "The app encourages me to accomplish tasks and make progress," Item 6: "The app supports me to overcome challenges"). Items are rated on a 7-point Likert-type scale from 1 (*strongly agree*) to 7 (*strongly disagree*). Internal consistency reliabilities computed for the DWAI are included in the Results section.

Social desirability.—The Socially Desirable Response Set (SRDS-5; Hays et al., 1989) was used in Study 1 to assess social desirability. Items reflect common but socially desirable or undesirable behavior (e.g., "There have been occasions when I took advantage of someone"). Items are rated on a 5-point Likert-type scale ranging from 1 (*definitely true*) to 5 (*definitely false*) in relation to how much statements are true or false for a given respondent. Items are scored as 1 or 0, with 1 assigned when an individual indicates the most socially desirable response (e.g., definitely false for a socially undesirable item). A total score is computed by summing across all five items. The measure has shown adequate reliability (internal consistency, test-retest; Hays et al., 1989) and validity (convergent, discriminant; Pechorro et al., 2016). Internal consistency was adequate in Study 1 (α = .71). As social desirability is a construct theoretically unrelated to alliance, it was used to assess discriminant validity.

Psychological distress.—Symptoms of depression and anxiety were assessed using the Patient-Reported Outcomes Measurement Information System (PROMIS) Depression and Anxiety measures (Pilkonis et al., 2011). Depression and anxiety were measured in Study 1 and at both baseline and post-treatment in Study 2. These measures have shown strong convergent validity with legacy measures of depression and anxiety (Choi et al., 2014;

Schalet et al., 2014), with short forms created using item response theory (Pilkonis et al., 2011). Study 1 used the four-item versions (4a) and Study 2 used the computer adaptive test (CAT) versions (v1.0) with the actual length varying depending on participant responses (Pilkonis et al., 2011). Items reflect symptoms of depression (e.g., "I felt worthless") and anxiety (e.g., "I felt fearful") and are rated based on the past 7 days on a 5-point Likert-type scale ranging from 1 (*never*) to 5 (*always*). A total score is computed by summing across items for the fixed length form. The CAT version yields a T-score (i.e., population mean = 50, *SD* = 10). Internal consistency reliability was adequate Study 1 (αs = .93 and .90, for depression and anxiety, respectively) and cannot be computed for the Study 2 CAT version.

Psychological stress was also assessed in Study 2 at baseline and post-treatment using the 10-item Perceived Stress Scale (Cohen & Williamson, 1988). This widely used measure of psychological stress assesses experiences in the past month (e.g., "How often have you felt that you were unable to control the important things in your life?"). Items are rated on a 5-point Likert-type scale from 1 (*never*) to 5 (*very often*). The 10-item version has shown acceptable psychometric properties including evidence for convergent and discriminant validity (Roberti et al., 2006). A total score was computed by summing across all items. Internal consistency was adequate in Study 2 (α = .85).

Based on high correlations between measures of depression, anxiety, and stress in previous work (Goldberg et al., 2020b), the preregistered data analytic plan for the randomized controlled trial in Study 2 involved the creation of a composite psychological distress variable. To do so, total scores on the three psychological symptom measures were z-transformed and then averaged. For consistency, a psychological distress composite was also created for Study 1 in the same way, although this was based on only depression and anxiety (stress was not assessed). As psychological distress is theoretically unrelated to alliance (Flückiger et al., 2020), it was used to assess discriminant validity. Change in psychological distress was used to assess predictive validity in Study 2.

Treatment preference.—Prior to randomization, participants in Study 2 indicated their preference for HMP ("How much would you like to receive the Healthy Minds Program app?") and waitlist control ("How much would you like to be in the waitlist control?"). Ratings were made on a 7-point Likert-type scale from 1 (*not at all*) to 7 (*a great deal*). As separate items were used to assess preference for each condition and participants could provide high or low ratings for either or both conditions, we theorized that preferring the waitlist would not necessarily be related to later alliance with the HMP app. Therefore, preference for waitlist was used to assess discriminant validity. Based on the notion that participants preferring the HMP app condition at baseline would be more likely to experience agreement on the tasks and goals of HMP (i.e., indicating greater interest and/or openness to the HMP app content), preference for the HMP app was used to assess convergent validity.

App utilization.—Utilization was assessed differently across the two studies. In Study 1, participants indicated whether they used their most used smartphone-based meditation app daily, weekly, monthly, several times per year, or never. This item was dichotomized into regular use (i.e., daily or weekly use) or non-regular use (i.e., monthly, several times

per year, or never). In Study 2, objective usage data was gathered through the HMP app. Utilization was operationalized as the total number of days during the four-week study period on which an individual used the app (i.e., days on which a participant completed an activity within the HMP app). Utilization was used to assess predictive validity.

Perceived app effectiveness.—A single item assessed perceived effectiveness of participants' most used smartphone-based meditation app in Study 1. The item ("How effective have you found this app?") was rated on a 6-point Likert-type scale from 1 (*not at all effective*) to 6 (*very effective*). Perceived effectiveness was used to assess convergent validity.

Analyses

As the DWAI has not previously been evaluated, we first sought to establish the measure's factor structure. We first conducted exploratory factor analysis using data from Study 1. Based on high inter-correlations between alliance dimensions observed in previous studies (e.g., Tracey & Kokotovic, 1989), models with an oblique (promax) rotation were run using the 'factanal' function in R (R Core Team, 2018). We aimed to determine the ideal number of factors that provided simple structure (i.e., items loading highly on only one factor; Thompson, 2004). In addition, we examined a scree plot and associated eigenvalues derived using the 'eigen' function in R and conducted a parallel analysis which is a preferred method for determining the number of factors to retain (Thompson, 2004) using the 'fa.parallel' function in the 'psych' package in R (Revelle, 2020).

Having determined the appropriate number of factors and the corresponding item factor loadings, we conducted a series of confirmatory factor analyses (CFA) using DWAI data from Study 2. The primary purpose of these analyses was to see if the structure from Study 1 replicated in Study 2. Additionally, because Study 2 involves longitudinal data, we could evaluate whether items were related to one another over time above and beyond the correlation among the latent factors over time. Model 1 was a longitudinal CFA that used the preferred factor structure from Study 1 at each of the four DWAI administrations (Weeks 1, 2, 3, and 4 of the study). We included a covariance between the latent variables across time. Thus, Model 1 assumes that the only relationship between items both within- and between-time points was at the latent-variable level. Model 2 added a residual correlation between the same item across time points (e.g., item 1 at Week 1 correlated with item 1 at Week 2; item 2 at Week 1 with item 2 at Week 2; and so on). Model 2 constrained the correlations to be equal across time on a per item basis (i.e., six total correlations, one for each item). Model 3 was identical to Model 2 except that correlations were freely estimated. Finally, Model 4 included an autoregression correlation structure, wherein correlations decay as the time points get further apart.

Overall fit of the CFA models was evaluated based on the comparative fit index (CFI) and root mean square error of approximation (RMSEA) using standard indices (i.e., CFI $\geq .95$, RMSEA $\leq .06$; Thompson, 2004). Fit was compared across models using the Aikake Information Criterion (AIC) and the Bayesian Information Criterion (BIC). CFA was

conducted using Mplus Version 8.1 (Muthén & Muthén, 2017) and the 'MplusAutomation' package in R (Hallquist & Wiley, 2018).

Reliability was assessed in two ways. In both Study 1 and Study 2, we calculated internal consistency reliability based on Cronbach's α. As Study 2 involved repeated DWAI assessments, we calculated test-retest reliability between weeks as the intraclass correlation coefficient (ICC1; single raters, absolute agreement) using the 'ICC' function in the 'psych' package (Revelle, 2020).

Validity was assessed in several different ways. Discriminant validity was evaluated by examining associations between DWAI scores with theoretically unrelated constructs. These included social desirability (Study 1), baseline psychological distress (Study 1 and Study 2), and preference for waitlist (Study 2). Convergent validity was evaluated by examining associations between DWAI scores with theoretically related constructs. These included perceived app effectiveness (Study 1) and preference for the HMP condition (Study 2). Finally, tests of predictive validity examined associations between DWAI scores with app utilization and changes in psychological distress. Utilization was assessed by self-report in Study 1 and obtained objectively through the HMP app in Study 2. All validity tests relied on correlation coefficients (i.e., Pearson's *r*, which simplifies to a point biserial correlation when including a dichotomous variable like regular app use). Linear regression models were used to examine DWAI scores as a predictor of post-treatment psychological distress controlling for pre-treatment psychological distress in Study 2. A subsequent linear regression model added preference for the HMP condition as a predictor of post-treatment psychological distress to assess incremental validity. In order to evaluate whether the strength of the alliance-outcome association varied depending on when alliance was assessed, we fit models in Mplus with these time-specific associations constrained to be equivalent or unique (i.e., free to vary across DWAI assessment time points). Fit was compared using a log-likelihood ratio test. An exploratory analysis described below evaluated whether the predictive validity of the DWAI varied based on psychological distress at baseline (i.e., test of moderation), to allow comparisons with the broader alliance in psychotherapy literature which has focused on clinical samples (Horvath et al., 2011). With the exception of test-retest reliability and predictive validity tests, all models using Study 2 DWAI scores focused on baseline DWAI scores.

Results

Descriptive statistics for all non-demographic study variables are reported in Table 1. Inter-correlations between DWAI total scores and these measures are reported in Tables 2 and 3. In Study 1, DWAI scores were uncorrelated (*p*s > .050) with age ($r = -.08$), male gender ($r = -.04$), college education ($r = -.01$), and income below $50,000 ($r = -.10$). However, non-Hispanic White race/ethnicity was associated with lower DWAI scores ($r = -.15$, $p = .009$; means = 30.00 and 32.30, $SD = 6.75$ and 5.67, for non-Hispanic White participants and racial/ethnic minority participants, respectively). In Study 2, baseline DWAI scores were uncorrelated (*p*s > .050) with age ($r = -.02$), male gender ($r = -.05$), non-Hispanic White race/ethnicity ($r = .09$), college education ($r = .06$), and income below $50,000 ($r = -.02$). Skewness and kurtosis were acceptable for most measures (skewness < 2, kurtosis

< 7; Curran et al., 1996), with the exception of the measure of social desirability which was extremely kurtotic (zero-inflated). This measure was dichotomized for use in analyses by assigning a value of 1 to all participants who had scores ≥1. Figure 1 displays the distribution of DWAI scores for Study 1 and Study 2. As is typical for alliance ratings, the measure showed some evidence of ceiling effects (Tryon et al., 2008). This range restriction should attenuate rather than inflate effect size estimates (Cohen et al., 2003).

Exploratory and Confirmatory Factor Analysis

Factor loadings for one-, two-, and three-factor solutions with an oblique (promax) rotation are presented in Table 4. As can be seen, all items loaded (≥0.43) on the single factor in the one factor model. In the two-factor solution, multiple items (Items 2 and 4) demonstrated cross-loadings (i.e., loadings ≥0.30 on multiple factors). Cross-loading was observed in the three-factor solution along with a single-item factor. A single factor solution was also supported by examination of the scree plot which indicated the presence of only one component showing an eigenvalue > 1 (Supplemental Materials Figure 1) and through parallel analysis (Supplemental Materials Figure 2). Thus, we concluded that a one-factor solution was preferred.

Next, we conducted confirmatory factor analysis using DWAI data from Study 2. As shown in Table 5, four models were estimated and compared. Both Model 2 and Model 3 showed near acceptable fit based on the CFI and TLI (≥.95) and RMSEAs (≤.060; Thompson, 2004). As Model 2 and Model 3 are nested, a formal model comparison was conducted. This indicated superior fit for Model 3 (χ^2 [30] = 48.69, p = .017). However, the simpler Model 2 which constrained correlated residuals to be equal across time points fit the data best based on AIC and BIC. Thus, Model 2 was considered the final model. Factor loadings for both models are reported in Table 6, with residual correlations in Supplemental Materials Table 1. As expected, all items loaded highly on the single factor (loadings ≥ 0.49). Loadings were nearly identical across Model 2 and Model 3.

Reliability

Internal consistency reliability was high in Study 1 (α = .90) and at all four time points the DWAI was assessed in Study 2 (αs = .88, .92, .91, and .92, for Weeks 1, 2, 3, and 4, respectively).[1] Test-retest reliability ranged from *ICC*s = .55 to .68 in Study 2 (Table 3; *p*s < .001). Between Week 1 and Week 2, test-retest reliability was *ICC* = .65.

Validity

Discriminant.—The association between the DWAI with social desirability (Study 1 only), psychological distress, and preference for the waitlist condition (Study 2 only) was used to assess discriminant validity. The DWAI was not associated with either raw or dichotomized social desirability (*r*s = −.06 and −.02, respectively, *p*s > .050; Table 2). The DWAI was also not associated with psychological distress in either study (*r*s = −.11 and .10, *p* > .050, for

[1]As Study 2 was a longitudinal RCT, sample sizes for estimating internal consistency differed across time points (*n*s = 285, 264, 260, and 289, for Weeks 1, 2, 3, and 4, respectively).

Study 1 and baseline distress in Study 2, respectively). The DWAI was not associated with preference for the waitlist condition in Study 2 ($r = -.07$, $p > .050$).

Convergent.—The association between the DWAI with perceived app effectiveness (Study 1 only) and preference for the HMP condition (Study 2) was used to assess convergent validity. The DWAI was highly correlated with perceived app effectiveness ($r = .75$, $p < .001$). The DWAI was also correlated with preference for the HMP app condition ($r = .26$, $p < .001$).

Predictive.—We examined changes in psychological distress in Study 2 and app utilization in both Study 1 and Study 2 as assessments of predictive validity. The DWAI was associated with greater likelihood of regular app use (weekly or more frequent use) in Study 1 ($r = .42$, $p < .001$). Average days of HMP use in Study 2 was 11.92 over the 4 weeks of the study (range = 0 to 29; Table 1). All DWAI assessments (i.e., Weeks 1 to 4) were associated with greater HMP app use (rs = .17 to .22, $p < .01$). In support of incremental validity, the associations persisted when controlling for preference for the HMP app condition at baseline (rs = .16 to .23, ps < .050).

The key test of the value of the DWAI is arguably provided by its ability to predict changes in psychological distress over the course of an unguided smartphone-based intervention. As shown in Table 7, early DWAI scores (Weeks 1 and 2) did not predict post-treatment distress when controlling for baseline distress (βs = −.05 and −.08, respectively, ps > .050). In contrast, higher DWAI scores assessed in Weeks 3 and 4 were associated with lower post-treatment distress (βs = −.17 and −.13, respectively, ps < .01). These effects remained statistically significant when adjusting for multiple tests (i.e., four DWAI scores) using Benjamini and Hochberg's (1995) false discovery rate (p_{FDR}) adjustment method. In support of incremental validity, results were essentially unchanged when controlling for preference for the HMP app condition at baseline (Table 7). Constraining the association between DWAI and post-test distress (controlling for pre-test distress) to be uniform across DWAI assessment time points showed poorer fit than the unconstrained model in which these associations could vary freely (χ^2 [3] = 8.90, $p = .031$). This suggests that the strength of the alliance-outcome association was not uniform across all weeks of Study 2.

As the alliance-outcome literature is primarily based on clinical samples (i.e., those seeking psychotherapy; Horvath et al., 2011) and the current trial in Study 2 was conducted in the general population (i.e., school district employees), we conducted a set of exploratory analyses examining whether the strength of the alliance-outcome association varied by baseline psychological distress. Specifically, we added a DWAI by baseline distress interaction term to the linear regression models predicting post-treatment distress from DWAI scores controlling for baseline distress. The interaction term was not significant for DWAI scores assessed at Weeks 1, 2, or 3 (Bs = −0.016 to −0.0024, $p > .050$). However, the interaction between Week 4 DWAI scores and baseline distress was significant ($B = -0.021$, $p = .010$, $p_{FDR} = .041$). As shown in Figure 2, the association between Week 4 DWAI scores and change in psychological distress (residualized change scores) is stronger (more negative) for those with higher baseline distress. As a point of comparison, we examined the Week 4 alliance-outcome association when the sample was restricted to those with PROMIS

Depression or PROMIS Anxiety scores above the moderate or higher clinical cut-off (i.e., T ≥ 60; Choi et al., 2014; HealthMeasures, n. d.). In this subsample (n = 150), the standardized regression coefficient for Week 4 DWAI scores predicting post-treatment psychological distress was β = −0.24, 95% confidence interval [−0.38, −0.09], p = .002.

Discussion

The current study evaluated the psychometric properties of a brief, six-item measure designed to assess working alliance in the context of an unguided smartphone app (Henson et al., 2019). Across two samples – a cross-sectional online survey and the intervention arm of a randomized trial – we assessed the measure's factor structure and tested several aspects of reliability and validity. Overall, results suggested the DWAI may possess desirable psychometric properties. Exploratory factor analysis suggested a single factor solution, with all items loading highly. This is consistent with the factor structure found by Miragall et al. (2015) in their adapted WAI for use in virtual and augmented reality therapy. Longitudinal confirmatory factor analysis suggested constraining correlated residuals to be equal across time fit the data adequately well. Thus, although it appears important to allow items to correlate across time, allowing these correlations to vary across items or to decay across time points was not necessary. The DWAI showed high internal consistency and evidence for stability across time (i.e., test-retest reliability). Supporting the measure's discriminant validity, the DWAI was not correlated with social desirability, psychological distress, or preference for a waitlist condition. Supporting convergent validity, the DWAI was associated with perceived meditation app effectiveness as well as with preference for the meditation app condition in a randomized controlled trial.

Most importantly, DWAI scores predicted meaningful outcomes. In both studies, we found evidence that those reporting higher DWAI scores were more likely to self-report (Study 1) or behaviorally demonstrate (Study 2) higher app usage. The considerably higher effect size detected when predicting self-report app usage in Study 1 (r = .42) compared to objective usage in Study 2 (rs = .16 to .22) could be due to a host of differences between the studies (e.g., app being evaluated, retrospective vs. repeated DWAI assessment). It is also possible that the association seen in Study 1 is inflated due to self-report biases that can occur when estimating technology use (see Kaye et al., 2020). In addition, higher DWAI scores assessed in the latter half of a 4-week intervention study (Weeks 3 or 4) predicted larger pre-post reductions in psychological distress. In support of incremental validity, associations with days of app usage and changes in distress in Study 2 were essentially unchanged when controlling for baseline preference for the HMP app condition. This bolsters the possibility that the DWAI is assessing something beyond pre-treatment preferences and may therefore be more likely to reflect participants' experiences actually using the app.

Although statistically significant, associations between DWAI scores with changes in psychological distress were modest (βs = −.17 and −.13, for Week 3 and Week 4, respectively) and smaller than those typically observed in psychotherapy (i.e., r = .278; Flückiger et al., 2018). Results from an exploratory moderator test suggests the DWAI may more strongly predict outcomes for individuals experiencing higher symptoms at baseline. Indeed, when restricted to those with moderate or higher depression or anxiety symptoms at

baseline, the alliance-outcome association at Week 4 is closer to that found in psychotherapy ($\beta = -0.24$). Also in keeping with the broader psychotherapy literature (i.e., Flückiger et al., 2018), it appears that the alliance-outcome association strengthens when alliance is assessed later in an intervention and/or when alliance and outcome are assessed closer in time. While this may be a measurement artifact (e.g., measures assessed at the same point in time being mutually influenced by state effects), it may also be that the validity of the alliance increases as an individual has opportunities for further exposure to an intervention.

One unexpected finding worth noting was the small negative association between non-Hispanic White race/ethnicity and DWAI scores in Study 1 ($r = -.15$). This finding was not replicated in Study 2, which showed a very small and non-significant association in the opposite direction ($r = .09$). Assuming DWAI scores do not actually vary by race/ethnicity, it is possible the effect in Study 1 reflects a measurement issue (i.e., items are rated systematically differently by racial/ethnic minority vs. non-Hispanic White individuals). It is also possible the difference is substantively meaningful and that in the general population, racial/ethnic minorities who are exposed to meditation apps actually experience higher alliance than non-Hispanic White individuals. This would be a welcome possibility, particularly given that racial/ethnic minorities tend to engage with psychotherapy interventions at lower rates (Cook et al., 2014; Goldberg et al., 2020a) and to have less access to quality mental health care (Alegría et al., 2008). Given the potential for mHealth interventions to help reduce mental health inequity within historically underserved communities (Anderson-Lewis et al., 2018), it would be valuable to further investigate associations between DWAI scores and race/ethnicity in future studies.

Limitations and Future Directions

The promising psychometric characteristics aside, it is worth asking whether the construct measured by the DWAI is the same as that measured by the WAI and other alliance measures in traditional, face-to-face psychotherapy. DWAI assessed later in Study 2 indeed predicted changes in distress. However, in our view, additional conceptual and empirical work is needed to more fully characterize what precisely working alliance with a smartphone app means. How does the inherently relational nature of the alliance in psychotherapy translate to a relationship with technology? Unguided smartphone apps necessarily lack the interpersonal back-and-forth characteristic of traditional psychotherapy. Key therapeutic processes associated with the alliance (e.g., alliance rupture and repair; Eubanks et al., 2018) are presumably impossible in the absence of interaction. At once, apps can be designed to create a sense of connection between the user with the content and perhaps also with the app creators (e.g., by having a guide or narrator who leads users through app content). Thus, participants may indeed experience a genuine interpersonal connection of sorts.

Clarifying the conceptual overlap and divergence between the DWAI and alliance in psychotherapy is essential to avoiding the nominal fallacy of believing we have explained, measured, or understood the alliance with a smartphone app simply calling a measure “alliance” and showing it predicts outcome. This topic could be addressed in future qualitative studies seeking to more richly describe the relational elements of smartphone app-based interventions. Such efforts may naturally depart from traditional

conceptualizations of the alliance in psychotherapy, and may find, for example, that core elements of Bordin's (1979) model (e.g., bond) are less relevant or manifest differently in app-based interventions. Thus, development of future measures designed to assess alliance in mHealth need not use the WAI as a starting point. It will likewise be crucial to evaluate the degree to which the DWAI or other measures purported to assess alliance with technology demonstrate discriminant and incremental validity in relation to conceptually distinct constructs such as treatment satisfaction and treatment expectancy (Kirsch et al., 2018; Tetzlaff et al., 2005). The extremely high correlation between DWAI scores with perceived app effectiveness in Study 1 (r = .75) highlights the possibility that assessment of the DWAI is strongly linked or perhaps confounded with other constructs drawn from the nomological network of alliance.

It would also be valuable to take up in future research the decades-old debate regarding the causal direction of the alliance-outcome association (DeRubeis & Feeley, 1990; Tang & DeRubeis, 1999). Our study, like those included in Flückiger et al.'s (2018) meta-analysis, merely shows an alliance-outcome *association*. While it could be worthwhile examining the ways in which alliance and outcome inter-relate longitudinally (e.g., using cross-lagged panel models; Falkenström et al., 2013), mHealth technology may allow a more satisfying method for addressing this question. It may be logistically feasible and perhaps ethically permissible to manipulate alliance within the context of an unguided smartphone app. For example, one may randomly assign individuals to low or high alliance versions of a smartphone app (e.g., with or without content designed specifically to increase alliance). The HMP app investigated in Study 2 includes content designed to recreate a "guided" experience even within the unguided app format (e.g., recorded meditation instructions in which a narrator speaks directly to app users; Goldberg et al., 2020b). If outcomes are assessed more intensively, one could readily examine the impact of shorter-term alliance manipulations (e.g., receiving meditation practice instructions designed to augment a sense of connection with the app guide versus instructions lacking such content) on more proximal measures (e.g., mood, app utilization).

As the alliance showed consistent relationships with usage even when alliance was assessed early in the intervention, it may serve as a valuable predictor of treatment dropout and disengagement. mHealth interventions show notoriously low retention (Christensen et al., 2009; Eysenbach, 2005; Linardon & Fuller-Tyszkiewicz, 2020; Pratap et al., 2020). Alliance scores might therefore be used as a "canary," alerting of potential disengagement when there may still be time to re-engage users. Future studies could test this possibility using alliance scores to implement adaptive trial designs (e.g., sequential multiple assignment randomized trials [SMART]; Collins et al., 2007). An individual reporting low alliance scores could receive an alliance-boosting intervention component (e.g., encouraging text message, motivation enhancing content) which ultimately could be built into highly responsive mHealth interventions.

This study has several important limitations that are worth noting. First, we evaluated only one of several potential alliance measures that have been proposed for use in an mHealth context. It is entirely possible that other measures (e.g., Berry et al., 2018; Herrero et al., 2020; Miragall et al., 2015) may have performed as well or better than the DWAI. Second,

the DWAI only includes six items, which may have limited our ability to reliably detect more complex factor structures. Third, we did not include measures of some key constructs that will be important to examine to further evaluate discriminant validity (e.g., system usability). Fourth, although Study 1 included a fairly diverse sample in terms of gender, race/ethnicity, age, education, and income, Study 2 was more racially/ethnically homogenous and predominantly female. The fact that race/ethnicity was associated with DWAI scores in Study 1 highlights the possibility that this construct may vary across demographic groups. Thus, future studies in highly diverse sample are warranted. Fifth, both studies focused exclusively on smartphone-based meditation apps. While meditation apps are representative of the majority of mental health app use (Wasil et al., 2020), future studies should examine DWAI or other alliance measure within other kinds of smartphone apps.

Conclusion

Digital technology is changing many aspects of human life and has the potential to revolutionize health care as well (Torous et al., 2015). However, for mHealth interventions to reach their potential and in order to maximize the acceptability and efficacy of these approaches, it is vital to more deeply understand the psychological processes at play. The working alliance has proven to be a key ingredient across diverse psychotherapeutic modalities (Flückiger et al., 2018). Results from the current study support the notion that a digital corollary of the alliance exists within the context of an unguided smartphone app and can be reliably and validly measured using a brief self-report instrument. Future research using the DWAI and other measures designed to capture users' subjective experience with mHealth technology may be crucial for maximizing the public health impact of these tools.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We include data from an online survey that was registered at the Open Science Framework (https://osf.io/4h86s/?view_only=0e5d7ad85f87468ea40e047b3cf7c795). We include data from a randomized controlled trial that was registered at ClinicalTrials.gov (NCT04426318) and through the Open Science Framework (https://osf.io/eqgt7). Data from both projects are available by request. This research was supported by the National Center for Complementary and Integrative Health Grant K23AT010879 (Simon B. Goldberg), the National Institute of Mental Health Grant R01MH43454 (Richard J. Davidson), the Clinical and Translational Science Award program through the NIH National Center for Advancing Translational Sciences Grant UL1TR002373, the Chan Zuckerberg Initiative Grant 2020-218037 (Richard J. Davidson), a National Academy of Education / Spencer Postdoctoral Fellowship (Matthew J. Hirshberg), and with funding from the Wisconsin Center for Education Research (Simon B. Goldberg). Support for this research was also provided by generous donors to the School of Education of the University of Wisconsin-Madison, by the Graduate School through support from the Wisconsin Alumni Research Foundation (Kevin M. Riordan), and by the University of Wisconsin - Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation (Simon B. Goldberg). Richard J. Davidson is the founder, president, and serves on the board of directors for the nonprofit organization, Healthy Minds Innovations, Inc.

References

Aboujaoude E, Salame W, & Naim L (2015). Telemental health: a status update. World Psychiatry, 14(2), 223–230. [PubMed: 26043340]

Alegría M, Chatterji P, Wells K, Cao Z, Chen C, Takeuchi D, et al. (2008). Disparity in depression among racial and ethnic minority populations in the United States. Psychiatric Services, 59(11), 1264–1272. [PubMed: 18971402]

Anderson-Lewis C, Darville G, Mercado RE, Howell S, & Di Maggio S (2018). mHealth technology use and implications in historically underserved and minority populations in the United States: systematic literature review. JMIR mHealth and uHealth, 6(6), e128. [PubMed: 29914860]

Baumel A, & Kane JM (2018). Examining predictors of real-world user engagement with self-guided eHealth interventions: analysis of mobile apps and websites using a novel dataset. Journal of Medical Internet Research, 20(12), e11491. [PubMed: 30552077]

Baumel A, Muench F, Edan S, & Kane JM (2019). Objective user engagement with mental health apps: systematic search and panel-based usage analysis. Journal of Medical Internet Research, 21(9), e14567. [PubMed: 31573916]

Benjamini Y, & Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, 57(1), 289–300.

Berger T (2017). The therapeutic alliance in internet interventions: A narrative review and suggestions for future research. Psychotherapy Research, 27(5), 511–524. [PubMed: 26732852]

Berry K, Salter A, Morris R, James S, & Bucci S (2018). Assessing therapeutic alliance in the context of mHealth interventions for mental health problems: Development of the mobile Agnew relationship measure (mARM) questionnaire. Journal of Medical Internet Research, 20(4), e90. [PubMed: 29674307]

Bordin ES (1979). The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy: Theory, Research and Practice, 16(3), 252–260.

Choi SW, Schalet B, Cook KF, & Cella D (2014). Establishing a common metric for depressive symptoms: Linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. Psychological Assessment, 26(2), 513–527. [PubMed: 24548149]

Christensen H, Griffiths KM, & Farrer L (2009). Adherence in internet interventions for anxiety and depression: Systematic review. Journal of Medical Internet Research, 11(2), e13. [PubMed: 19403466]

Cohen J, & Torous J (2019). The potential of object-relations theory for improving engagement with health apps. JAMA, 322(22), 2169–2170. [PubMed: 31725854]

Cohen S, & Williamson G (1988). Perceived stress in a probability sample of the United States. In Spacapan S & Oskamp S (Eds.), The social psychology of health: Claremont Symposium on Applied Social Psychology (pp. 31–67). Newbury Park, CA: Sage.

Collins LM, Murphy SA, & Strecher V (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): New methods for more potent eHealth interventions. American Journal of Preventive Medicine, 32(5), S112–S118. [PubMed: 17466815]

Cook BL, Zuvekas SH, Carson N, Wayne GF, Vesper A, & McGuire TG (2014). Assessing racial/ethnic disparities in treatment across episodes of mental health care. Health Services Research, 49(1), 206–229. [PubMed: 23855750]

Dahl CJ, Wilson-Mendenhall CD, & Davidson RJ (2020). The plasticity of well-being: A training-based framework for the cultivation of human flourishing. Proceedings of the National Academy of Sciences, 117(51), 32197–32206.

DeRubeis RJ & Feeley M (1990). Determinants of change in cognitive therapy for depression. Cognitive Therapy and Research, 14, 469–482.

Eubanks CF, Muran JC, & Safran JD (2018). Alliance rupture repair: A meta-analysis. Psychotherapy, 55(4), 508–519. doi: 10.1037/pst0000185 [PubMed: 30335462]

Eysenbach G (2005). The law of attrition. Journal of Medical Internet Research, 7(1), e11. [PubMed: 15829473]

Falkenström F, Granström F, & Holmqvist R (2013). Therapeutic alliance predicts symptomatic improvement session by session. Journal of Counseling Psychology, 60(3), 317–328. doi: 10.1037/a0032258 [PubMed: 23506511]

Falkenström F, Hatcher RL, & Holmqvist R (2015a). Confirmatory factor analysis of the patient version of the Working Alliance Inventory–Short Form Revised. Assessment, 22(5), 581–593. [PubMed: 25271007]

Falkenström F, Hatcher RL, Skjulsvik T, Larsson MH, & Holmqvist R (2015b). Development and validation of a 6-item working alliance questionnaire for repeated administrations during psychotherapy. Psychological Assessment, 27(1), 169–183. [PubMed: 25346997]

Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, & Sarris J (2017a). The efficacy of smartphone-based mental health interventions for depressive symptoms: A meta-analysis of randomized controlled trials. World Psychiatry, 16(3), 287–298. [PubMed: 28941113]

Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, & Sarris J (2017b). Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. Journal of Affective Disorders, 218, 15–22. doi: 10.1016/j.jad.2017.04.046 [PubMed: 28456072]

Flückiger C, Del Re AC, Wampold BE, & Horvath AO (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. Psychotherapy, 55(4), 316–340. doi: 10.1037/pst0000172 [PubMed: 29792475]

Flückiger C, Del Re AC, Wlodasch D, Horvath AO, Solomonov N, & Wampold BE (2020). Assessing the alliance–outcome association adjusted for patient characteristics and treatment processes: A meta-analytic summary of direct comparisons. Journal of Counseling Psychology, 67(6), 706–711. 10.1037/cou0000424 [PubMed: 32212755]

Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, … & Kessler D (2015). Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ, 351, h5627. [PubMed: 26559241]

Goldberg SB, Fortney JC, Chen JA, Young BA, Lehavot K, & Simpson TL (2020a). Military service and military health care coverage are associated with reduced racial disparities in time to mental health treatment initiation. Administration and Policy in Mental Health and Mental Health Services Research, 47, 555–568. doi: 10.1007/s10488-020-01017-2 [PubMed: 31989399]

Goldberg SB, Imhoff-Smith T, Bolt DM, Wilson-Mendenhall CD, Dahl CJ, Davidson RJ, & Rosenkranz MA (2020b). Testing a multi-component, self-guided, smartphone-based meditation app: three-armed randomized controlled trial. JMIR Mental Health, 7(11), e23825. doi: 10.2196/23825 [PubMed: 33245288]

Hallquist MN, & Wiley JF (2018). MplusAutomation: an R package for facilitating large-scale latent variable analyses in M plus. Structural Equation Modeling, 25(4), 621–638. [PubMed: 30083048]

Hays RD, Hayashi T, & Stewart AL (1989). A five-item measure of socially desirable response set. Educational and Psychological Measurement, 49(3), 629–636.

HealthMeasures. (n. d.). PROMIS Score Cut Points. Retrieved on July 5, 2020 from: https://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/promis-score-cut-points

Henson P, Wisniewski H, Hollis C, Keshavan M, & Torous J (2019). Digital mental health apps and the therapeutic alliance: Initial review. BJPsych Open, 5(1).

Herrero R, Vara M, Miragall M, Botella C, García-Palacios A, Riper H, … & Baños RM (2020). Working Alliance Inventory for Online Interventions-Short Form (WAI-TECH-SF): The role of the therapeutic alliance between patient and online program in therapeutic outcomes. International Journal of Environmental Research and Public Health, 17(17), 6169.

Horvath AO, Del Re AC, Flűckiger C, & Symonds D (2011). Alliance in individual psychotherapy. Psychotherapy, 48(1), 9–16. doi: 10.1037/a0022186 [PubMed: 21401269]

Horvath A & Greenberg L (1989). Development and validation of the working alliance inventory. Journal of Counseling Psychology, 36(2), 223–233. doi: 10.1037/0022-0167.36.2.223

Horvath A & Symonds B (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. Journal of Counseling Psychology, 38(2), 139–149.

Kaye LK, Orben A, A Ellis D, C Hunter S, & Houghton S (2020). The conceptual and methodological mayhem of “screen time”. International Journal of Environmental Research and Public Health, 17(10), 3661.

Kiluk BD, Serafini K, Frankforter T, Nich C, & Carroll KM (2014). Only connect: The working alliance in computer-based cognitive behavioral therapy. Behaviour Research and Therapy, 63, 139–146. [PubMed: 25461789]

Kirsch V, Keller F, Tutus D, & Goldbeck L (2018). Treatment expectancy, working alliance, and outcome of Trauma-Focused Cognitive Behavioral Therapy with children and adolescents. Child and Adolescent Psychiatry and Mental Health, 12(1), 16. [PubMed: 29515647]

Linardon J, Cuijpers P, Carlbring P, Messer M, & Fuller-Tyszkiewicz M (2019). The efficacy of app-supported smartphone interventions for mental health problems: A meta- analysis of randomized controlled trials. World Psychiatry, 18(3), 325–336. [PubMed: 31496095]

Linardon J, & Fuller-Tyszkiewicz M (2020). Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. Journal of Consulting and Clinical Psychology, 88(1), p. 1–13. doi:10.1037/ccp0000459 [PubMed: 31697093]

Liu S, Yang L, Zhang C, Xiang YT, Liu Z, Hu S, & Zhang B (2020). Online mental health services in China during the COVID-19 outbreak. The Lancet Psychiatry, 7(4), e17–e18. 10.1016/S2215-0366(20)30077-8 [PubMed: 32085841]

Miloff A, Carlbring P, Hamilton W, Andersson G, Reuterskiöld L, & Lindner P (2020). Measuring Alliance Toward Embodied Virtual Therapists in the Era of Automated Treatments With the Virtual Therapist Alliance Scale (VTAS): Development and Psychometric Evaluation. Journal of Medical Internet Research, 22(3), e16660. [PubMed: 32207690]

Miragall M, Baños RM, Cebolla A, & Botella C (2015). Working alliance inventory applied to virtual and augmented reality (WAI-VAR): psychometrics and therapeutic outcomes. Frontiers in Psychology, 6, 1531. [PubMed: 26500589]

Muthén LK, & Muthén BO (1998–2017). Mplus User's Guide. Eight Edition. Los Angeles, CA: Muthén & Muthén.

Osenbach JE, O'Brien KM, Mishkind M, & Smolenski DJ (2013). Synchronous telehealth technologies in psychotherapy for depression: A meta-analysis. Depression and Anxiety, 30(11), 1058–1067. [PubMed: 23922191]

Pechorro P, Ayala-Nunes L, Oliveira JP, Nunes C, & Goncalves RA (2016). Psychometric properties of the Socially Desirable Response Set-5 among incarcerated male and female juvenile offenders. International Journal of Law and Psychiatry, 49, 17–21. [PubMed: 27210577]

Peer E, Brandimarte L, Samat S, & Acquisti A (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. Journal of Experimental Social Psychology, 70, 153–163.

Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D, & PROMIS Cooperative Group. (2011). Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS): Depression, anxiety, and anger. Assessment, 18(3), 263–283. [PubMed: 21697139]

Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, … & Mangravite L (2020). Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. npj digital medicine, 3(1), 1–10. [PubMed: 31934645]

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Revelle W (2020). psych: Procedures for personality and psychological research. R package version 2.0.8, http://CRAN.R-project.org/package=psych

Roberti JW, Harrington LN, & Storch EA (2006). Further psychometric support for the 10-item version of the perceived stress scale. Journal of College Counseling, 9(2), 135–147.

Schalet BD, Cook KF, Choi SW, & Cella D (2014). Establishing a common metric for self-reported anxiety: Linking the MASQ, PANAS, and GAD-7 to PROMIS Anxiety. Journal of Anxiety Disorders, 28(1), 88–96. [PubMed: 24508596]

Tang TZ, & DeRubeis RJ (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. Journal of Consulting and Clinical Psychology, 67(6), 894. [PubMed: 10596511]

Tetzlaff BT, Kahn JH, Godley SH, Godley MD, Diamond GS, & Funk RR (2005). Working alliance, treatment satisfaction, and patterns of posttreatment use among adolescent substance users. Psychology of Addictive Behaviors, 19(2), 199–207. 10.1037/0893-164X.19.2.199 [PubMed: 16011391]

Thompson B (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. Washington, DC: American Psychological Association.

Torous J, Myrick KJ, Rauseo-Ricupero N, & Firth J (2020). Digital mental health and COVID-19: Using technology today to accelerate the curve on access and quality tomorrow. JMIR Mental Health, 7(3), e18848. [PubMed: 32213476]

Torous J, Staples P, & Onnela JP (2015). Realizing the potential of mobile mental health: new methods for new data in psychiatry. Current Psychiatry Reports, 17(8), 61.

Tracey T & Kokotovic A (1989). Factor structure of the working alliance inventory. Psychological Assessment, 1(3), 207–210.

Wasil AR, Gillespie S, Patel R, Petre A, Venturo-Conerly KE, Shingleton RM, … & DeRubeis RJ (2020). Reassessing evidence-based content in popular smartphone apps for depression and anxiety: Developing and applying user-adjusted analyses. Journal of Consulting and Clinical Psychology, 88(11), 983. [PubMed: 32881542]

Wehmann E, Köhnen M, Härter M, & Liebherz S (2020). Therapeutic Alliance in Technology-Based Interventions for the Treatment of Depression: Systematic Review. Journal of Medical Internet Research, 22(6), e17195. [PubMed: 32525484]

Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, & Ebert DD (2019). Standalone smartphone apps for mental health—a systematic review and meta-analysis. npj digital medicine, 2(1), 1–10. 10.1038/s41746-019-0188-8 [PubMed: 31304351]

Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, & Smith AC (2020). The role of telehealth in reducing the mental health burden from COVID-19. Telemedicine and e-Health, 26(4), 377–379. [PubMed: 32202977]

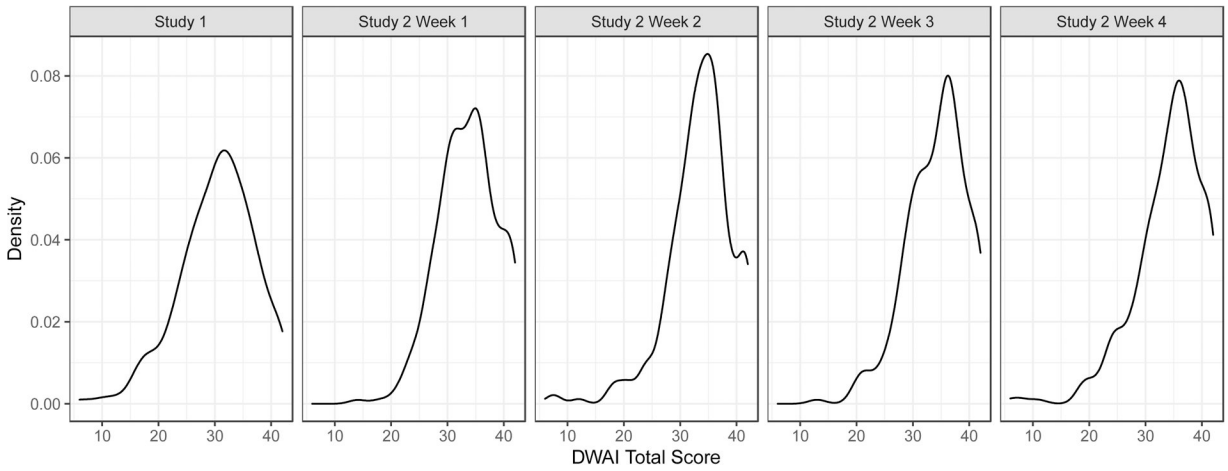

Figure 1.

Density distributions for Digital Working Alliance Inventory (DWAI) scores from Study 1 (*n* = 290) and at Weeks 1, 2, 3, and 4 in Study 2 (*n*s = 262 to 290).

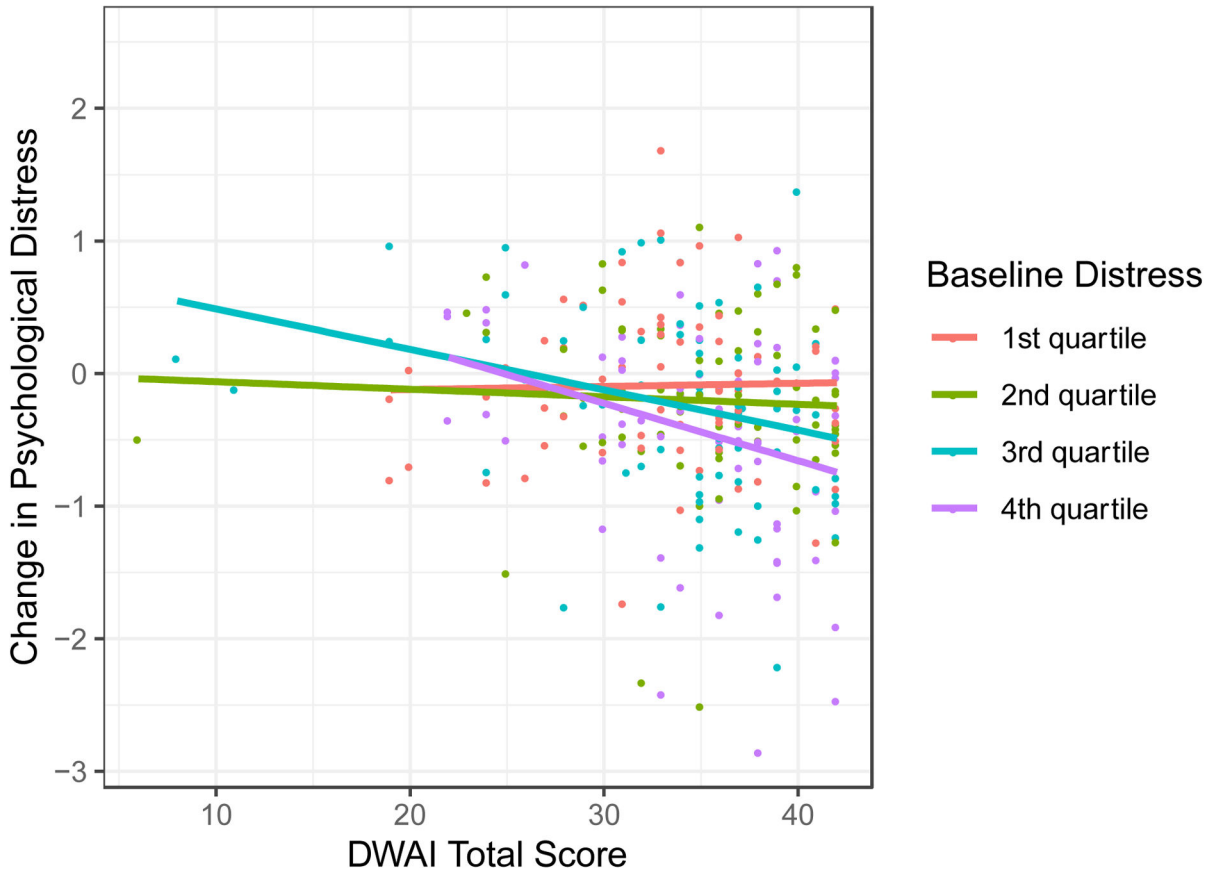

Figure 2.

Week 4 Digital Working Alliance Inventory (DWAI) scores predicting pre-post residualized change in psychological distress in Study 2. Smaller (more negative) residualized change scores indicate pre-post improvement (i.e., reductions) in psychological distress. The association is stronger (more negative) for those with higher distress at baseline. Baseline distress was split into four quartiles for plotting purposes. $n = 285$.

Table 1.

Descriptive statistics for Study 1 and Study

Sample	Variable	*n*	Mean	SD	Min	Max	Skew	Kurtosis
Study 1	DWAI	290	30.58	6.56	6	42	−0.52	0.48
	Social desirability raw	290	0.17	0.59	0	5	5.23	33.79
	Social desirability dich	290	0.11	0.31	0	1	2.47	4.14
	Distress	290	0.17	0.92	−1.16	2.76	0.30	−0.76
	Regular app use	290	0.36	0.48	0	1	0.60	−1.64
	App effectiveness	290	4.22	1.29	1	6	−0.59	−0.18
Study 2	Week 1 DWAI	286	33.56	5.19	14	42	−0.33	−0.02
	Week 2 DWAI	265	33.5	5.70	7	42	−1.21	3.16
	Week 3 DWAI	261	34.24	5.25	13	42	−0.68	0.53
	Week 4 DWAI	290	34.13	6.00	6	42	−1.23	2.59
	Baseline distress	306	−0.02	0.87	−2.96	2.03	−0.28	−0.07
	Week 4 distress	290	−0.67	0.89	−3.34	1.91	−0.14	0.28
	Prefer HMP	309	5.32	1.36	1	7	−0.36	−0.45
	Prefer waitlist	308	3.67	1.34	1	7	−0.10	0.25
	Days of HMP use	314	11.92	8.89	0	29	0.04	−1.34

Note: DWAI = Digital Working Alliance Inventory; Social desirability = Socially Desirable Response Set – 5 in raw score units or dichotomized (dich) (0 and ≥1); Distress = composite of Patient-Reported Outcome Information System Depression and Anxiety (Study 1), with Perceived Stress Scale also combined in Study 2; Regular app use = daily or weekly app use; HMP = Healthy Minds Program app.

Table 2.

Intercorrelations between Digital Working Alliance Inventory scores and Study 1 variables

	1	2	3	4	5
1. DWAI					
2. Social desirability raw	−0.06				
3. Social desirability dich	−0.02	0.80***			
4. Distress	−0.11	0.12 *	0.12 *		
5. Regular app use	0.42***	0.00	−0.01	−0.07	
6. App effectiveness	0.75***	−0.05	−0.02	−0.20***	0.49***

Note: DWAI = Digital Working Alliance Inventory; Social desirability = Socially Desirable Response Set – 5 in raw score units or dichotomized (dich) (0 and ≥1); Distress = composite of Patient-Reported Outcome Information System Depression and Anxiety (Study 1), with Perceived Stress Scale also combined in Study 2; Regular app use = daily or weekly app use (coded as 1 = regular app use, 0 = not regular app use); HMP = Healthy Minds Program app. $n = 290$.

* $p < .05$,

** $p < .01$,

** $p < .001$

Assessment. Author manuscript; available in PMC 2023 September 01.

Table 3.

Intercorrelations between Digital Working Alliance Inventory scores and Study 2 variables

	1	2	3	4	5	6	7	8
1. Week 1 DWAI								
2. Week 2 DWAI	0.65***							
3. Week 3 DWAI	0.65***	0.65***						
4. Week 4 DWAI	0.55***	0.61***	0.68***					
5. Baseline distress	0.10	−0.02	−0.01	0.09				
6. Week 4 distress	0.00	−0.11	−0.18**	−0.07	0.63***			
7. Prefer HMP	0.26***	0.08	0.20***	0.24***	0.27***	0.15*		
8. Prefer waitlist	−0.07	0.08	−0.07	−0.03	−0.17**	0.01	−0.34***	
9. Days of HMP use	0.21***	0.22***	0.16*	0.17**	0.03	−0.03	0.09	−0.09

Note: Associations between DWAI scores across weeks (i.e., test-retest reliability) are intraclass correlation coefficients (ICC1) while all other values are Pearson's *r*. DWAI = Digital Working Alliance Inventory; HMP = Healthy Minds Program app. *n*s = 236 to 309.

*$p < .05$,

**$p < .01$,

***$p < .001$

Table 4.

Results of exploratory factor analysis with oblique (Promax) rotation in Study 1

	One Factor	Two Factor		Three Factor		
Item	**Factor 1**	**Factor 1**	**Factor 2**	**Factor 1**	**Factor 2**	**Factor 3**
1	0.864	0.117	0.845	0.227	0.403	0.285
2	0.861	0.485	0.412	0.124	0.010	0.894
3	0.784	0.810	0.005	0.541	0.372	−0.085
4	0.845	0.568	0.310	0.017	0.928	−0.012
5	0.434	0.493	−0.045	0.517	−0.057	−0.002
6	0.807	0.814	0.025	0.877	−0.063	0.059

Note: Values indicate factor loadings. $n = 290$.

Table 5.

Model fit for confirmatory factor analysis with repeated Digital Working Alliance Inventory assessments in Study 2

Model	# Param	χ^2	*df*	CFI	TLI	AIC	BIC	RMSEA	90% CI_{LB}	90% CI_{UB}
Model 1	78	1139.415	246	0.846	0.827	15338.712	15631.165	0.108	0.101	0.114
Model 2	84	453.779	240	0.963	0.958	14665.076	14980.025	0.053	0.046	0.061
Model 3	114	405.091	210	0.966	0.956	14676.388	15103.819	0.054	0.046	0.062
Model 4	66	837.234	258	0.900	0.893	15012.531	15259.991	0.085	0.078	0.091

Note: Model 1 = independent factor structure across time points; Model 2 = correlated residuals constrained to be equal across time points; Model 3 = correlated residuals not constrained; Model 4 = correlated residuals with autoregressive (AR1) covariance structure. # Param = number of parameters; CFI = comparative fit index; TLI = Tucker-Lewis fit index; AIC = Aikake information criterion; BIC = Bayesian information criterion; RMSEA = root mean square error of approximation; 90% CI = 90% confidence interval lower (LB) and upper (UB) bounds.

Table 6.

Standardized factor loadings for final confirmatory factor analysis models

		Model 2		Model 3	
Time	**Item**	**Estimate**	***SE***	**Estimate**	***SE***
Week 1	1	0.782	0.026	0.797	0.025
Week 1	2	0.801	0.024	0.791	0.025
Week 1	3	0.774	0.026	0.794	0.025
Week 1	4	0.867	0.019	0.869	0.019
Week 1	5	0.492	0.041	0.496	0.041
Week 1	6	0.790	0.025	0.797	0.025
Week 2	1	0.858	0.018	0.871	0.018
Week 2	2	0.859	0.018	0.859	0.018
Week 2	3	0.834	0.019	0.824	0.021
Week 2	4	0.895	0.015	0.895	0.015
Week 2	5	0.618	0.034	0.611	0.035
Week 2	6	0.835	0.020	0.836	0.021
Week 3	1	0.887	0.015	0.876	0.017
Week 3	2	0.868	0.018	0.882	0.017
Week 3	3	0.808	0.022	0.805	0.023
Week 3	4	0.856	0.019	0.853	0.019
Week 3	5	0.571	0.037	0.565	0.037
Week 3	6	0.801	0.024	0.796	0.025
Week 4	1	0.889	0.014	0.879	0.016
Week 4	2	0.887	0.014	0.886	0.015
Week 4	3	0.837	0.018	0.833	0.019
Week 4	4	0.901	0.013	0.902	0.014
Week 4	5	0.607	0.033	0.601	0.034
Week 4	6	0.818	0.021	0.819	0.021

Note: Results reported for Model 2 which included correlated residuals constrained to be equal across time points and Model 3 which allowed residuals to vary across time points. *SE* = standard error; *p*s < .001 for all factor loadings.

Table 7.

Results of linear regression models predicting post-treatment distress

Model	DWAI Week	β	$SE\ \beta$	95% CI_{LB}	95% CI_{UB}	p	p_{FDR}
Unadjusted	Week 1	−0.050	0.048	−0.145	0.045	.301	.301
Unadjusted	Week 2	−0.082	0.050	−0.181	0.018	.107	.143
Unadjusted	Week 3	−0.174	0.049	−0.270	−0.077	< .001	< .001
Unadjusted	Week 4	−0.131	0.046	−0.221	−0.041	.004	.008
Adjusted	Week 1	−0.046	0.050	−0.144	0.052	.358	.358
Adjusted	Week 2	−0.080	0.051	−0.180	0.020	.117	.156
Adjusted	Week 3	−0.178	0.050	−0.277	−0.079	< .001	< .001
Adjusted	Week 4	−0.136	0.047	−0.228	−0.043	.004	.008

Note: β are standardized regression coefficients for DWAI scores assessed at Weeks 1, 2, 3 or 4 as predictors of post-treatment distress controlling for pre-treatment distress. Adjusted models added baseline preference for the Healthy Minds Program (HMP) app condition as a covariate to assess incremental validity. *SE* = standard error; 95% CI = 95% confidence interval lower (LB) and upper (UB) bound; p = p-value for DWAI coefficient; pFDR = p-value adjusted for multiple tests using Benjamini and Hochberg's (1995) method.