

Original Paper

Ecological Momentary Assessment as a Measure of Intervention Change: Evaluation in 4 Digital Mental Health Trials

Christian A Webb^{1,2}, PhD; Lori M Hilt³, PhD; Caroline M Swords⁴, BA; Daniel M Bolt⁵, PhD; Hadar Fisher^{1,2}, PhD; Simon B Goldberg⁴, PhD

¹McLean Hospital, Belmont, MA, United States

²Harvard Medical School, Boston, MA, United States

³Department of Psychology, Lawrence University, Appleton, WI, United States

⁴Center for Healthy Minds and Department of Counseling Psychology, University of Wisconsin–Madison, Madison, WI, United States

⁵Department of Educational Psychology, University of Wisconsin–Madison, Madison, WI, United States

Corresponding Author:

Christian A Webb, PhD

McLean Hospital

115 Mill St

Belmont, MA, 02474

United States

Phone: 1 6178554429

Email: cwebb@mclean.harvard.edu

Abstract

Background: Ecological momentary assessment (EMA) is increasingly being incorporated into intervention studies to acquire a more fine-grained and ecologically valid assessment of change. The added utility of including relatively burdensome EMA measures in a clinical trial hinges on several psychometric assumptions, including that these measure are (1) reliable, (2) related to but not redundant with conventional self-report measures (convergent and discriminant validity), (3) sensitive to intervention-related change, and (4) associated with a clinically relevant criterion of improvement (criterion validity) above conventional self-report measures (incremental validity).

Objective: This study aimed to evaluate the reliability, validity, and sensitivity to change of conventional self-report versus EMA measures of rumination improvement.

Methods: We conducted a secondary analysis of data from 4 trials of app-based meditation interventions (N=412). Participants included adolescents (samples 1-2; n=232, 56.3%; aged 12-15 years), college students (sample 3; n=88, 21.4%; aged 18-21 years), and adults (sample 4; n=92, 22.3%; aged 18-80 years). All participants completed validated conventional self-report rumination measures (Children's Response Styles Questionnaire, Response Styles Questionnaire, or Perseverative Thinking Questionnaire) at baseline and the postintervention time point and EMA rumination assessments throughout 3- to 4-week intervention periods (mean compliance 72.2%-80.2%). We examined the reliability of conventional self-report and EMA measures, computed correlations between measurement approaches, compared rumination improvement using standardized effect sizes, and tested criterion validity by predicting depression symptom improvement.

Results: Reliability of conventional self-report rumination was high at baseline (Cronbach α =0.89-0.94) and the postintervention time point (Cronbach α =0.90-0.95) but decreased for residualized change scores (p = 0.71-0.90). EMA rumination showed high mean-level reliability (p =0.89-0.96) but substantially lower reliability for change scores (p =0.50-0.77). Conventional self-report and EMA measures showed medium correlations at individual time points (r =0.28-0.47; P <.01 in all cases) but nonsignificant correlations between change scores (Child and Adolescent Research in Emotion samples: r =0.03 and P =.66; Healthy Minds Program sample: r =0.20 and P =.06). Conventional self-report detected larger intervention effects than EMA measures (Cohen d =0.37 vs 0.14 for group differences; Cohen d =0.77 vs 0.17 for pretest-posttest change). Despite modest intercorrelations, both measurement approaches demonstrated incremental predictive validity for depression improvement. In the Child and Adolescent Research in Emotion samples, improvements in both conventional rumination (b =0.04; SE 0.017; P =.03) and EMA rumination (b =0.03; SE 0.014; P =.04) significantly predicted reduced depression symptoms until the 12-week follow-up. Similarly, in the Healthy Minds Program sample, both conventional rumination change (b =0.23; SE 0.060; P <.001) and EMA rumination change (b =0.20; SE 0.094; P =.04) predicted depression improvement.

Conclusions: Conventional self-report and EMA measures of rumination provide distinct and clinically meaningful information. When deciding to use EMA in intervention studies, researchers should carefully consider the psychometric properties of their measures and the precise construct they intend to capture.

(*J Med Internet Res* 2025;27:e69297) doi: [10.2196/69297](https://doi.org/10.2196/69297)

KEYWORDS

ecological momentary assessment; rumination; mindfulness; apps; reliability

Introduction

Background

An increasing number of intervention studies (particularly mobile mental health trials) are incorporating smartphone-delivered ecological momentary assessment (EMA) to measure symptom improvement [1-9]. Proponents of EMA argue that the reliance on commonly used retrospective self-report questionnaires may be convenient and cost-effective, but these measures often lack ecological validity and are contaminated by recall bias [2,10,11]. Specifically, in contrast to EMA surveys, which are deployed in daily life and ask about current or recent (eg, “since the last survey”) experiences, conventional self-report symptom questionnaires typically ask participants to recall and summarize their average levels of symptoms over a relatively lengthy period (eg, several weeks). For example, the commonly used Beck Depression Inventory [12] and Center for Epidemiologic Studies Depression Scale [13] ask participants to recall and summarize their experience of depressive symptoms over the previous 2 weeks and 1 week, respectively. It is worth noting that common clinician-administered symptom measures (eg, Hamilton Depression Rating Scale [14] or Children’s Depression Rating Scale [15]) are not immune to these issues as they are still based on participant self-report to the interviewer and ask about time frames that extend into the past. However, skillful clinician questioning may improve participant recall and reduce memory biases relative to conventional self-report questionnaires. At an even grosser level of temporal abstraction, trait or global self-report measures ask respondents to report on their *typical* tendencies or personal characteristics without specifying a recall time frame. For example, the Response Styles Questionnaire (RSQ) [16] and Children’s RSQ (CRSQ) [17] are commonly used to assess depressive rumination in adults and children, respectively. These measures ask participants what they “generally” or “usually” do when feeling sad.

Previous research reveals that retrospective reports are often biased. For example, studies show that both children [18,19] and adults [20-24] overestimate levels of symptoms on retrospective self-report measures relative to averaged EMA reports (for evidence of age moderating this effect, see the work by Neubauer et al [25] and Zurbriggen et al [26]). Studies have also shown that, when individuals report on their experiences (eg, pain during a medical procedure) via retrospective self-report, they often incorporate information about the most intense (peak) moment and how the experience ended (ie, the so-called peak-end bias) rather than equally weighting time points and simply averaging their momentary experiences over the entire reporting time frame [27-30]. This

“memory-experience gap” is due to the fact that momentary EMA requires introspection of current states (eg, emotions, thoughts, and behaviors in the here and now), whereas retrospective self-report taps into autobiographical (episodic) memory, which is limited by the extent to which past experiences and events have been accurately encoded and consolidated in memory [11,31,32]. Moreover, memory retrieval has been shown to be biased by one’s current state (eg, mood-congruent memory bias [33]). This is further complicated by the fact that individuals may shift from episodic memory to semantic memory (or a combination of the 2) when recall time frames are longer or remembering relevant details of past episodes becomes challenging [11,34]. Semantic knowledge includes more abstract beliefs or generalization about the self that are not tied to a specific time and place (eg, “I’m generally a happy person” or “I’m a worrier” [11,34,35]). One study indicated that the shift from relying on episodic memory in reporting past emotional well-being to drawing on both semantic and episodic retrieval strategies was in the 3- to 7-week range [36] (but see also the work by Walentynowicz et al [37]).

Collectively, these findings suggest that intervention researchers may benefit from relying less on convenient but often biased retrospective self-report measures and, instead, adopting EMA to measure symptom improvement. However, EMA does not necessarily provide an inherently more reliable and valid measure of intervention-related change. The repeated assessments involved in EMA protocols may introduce unique biases and careless responding (eg, random or stereotyped responding) among at least some participants [38]. For example, there is some evidence of an “initial elevation bias” (ie, an upward bias in scores on initial self-reports) when subjective reports are collected repeatedly, and this effect is more pronounced for negative mental states and physical symptoms than for positive states and behaviors [39]. However, a subsequent study found these effects to be inconsistent and minimal [40]. As another example, a recent study applied drift diffusion modeling to reaction time data from EMA affect ratings and found evidence of a shift over time in the cognitive processes that underlie survey responses. Specifically, the authors found pronounced changes in 2 central drift diffusion parameters across the repeated EMA affect ratings: increasing “drift rate” (reflecting faster processing of affective information over time) and decreased “boundary separation” (reflecting a decrease in how thoroughly participants process items and how cautiously they respond over time) [41,42]. In summary, at this stage, given conflicting findings, it is not entirely clear to what extent EMAs are more reliable and valid measures relative to conventional retrospective self-report instruments.

Objectives

The added utility of including relatively burdensome EMA measures of change in an intervention study hinges on several psychometric assumptions, including that these measures are (1) reliable, (2) related to but not redundant with conventional self-report measures of the same construct (convergent and discriminant validity), (3) sensitive to intervention-related change, and (4) associated with a clinically relevant criterion of patient improvement (criterion validity) above and beyond conventional self-report measures (incremental validity). In this study, we included data from 4 clinical trials (N=412) of app-delivered meditation training that incorporated both EMA and conventional self-report measures of improvement in rumination. We compared these measures with regard to their reliability, validity, and sensitivity to detecting change to help inform which measures to include in future intervention research.

It is important to note that we focused on EMA measures of improvement in rumination levels over time in this study. Beyond measuring change, EMA offers unique opportunities to examine psychological dynamics that may be markers of mental health and intervention effects and are impossible to capture using traditional self-report measures. Dynamic applications of EMA have gained considerable attention in recent years, with growing interest in network models (eg, examining how symptoms influence each other over time) [43], complex dynamic system approaches (eg, studying emotional inertia and variability as indicators of psychological flexibility) [44,45], and temporal mediation analyses examining how changes in putative mediators precede and predict subsequent changes in outcomes within individuals over time [46]. However, poor reliability in EMA can obscure and interfere with the ability to detect such dynamics as well, ultimately undermining the very advantages that make EMA valuable [47].

Methods

Participants

Two of the trials involved adolescent samples aged 12 to 15 years selected from the community for moderate to high levels of rumination. The 80 participants in sample 1 (mean age 14.01, SD 0.99 years; n=36, 45% girls; n=69, 86% White; n=3, 4% Hispanic; median household income US \$100,000-\$125,000) were recruited from the Appleton, Wisconsin, area in 2018 to 2019 for a single-arm trial investigating a mindfulness mobile app. In total, 88% (70/80) of the participants completed the 3-week trial (for more details, see the study by Hilt and Swords [1]). The 152 participants in sample 2 (mean age 13.71, SD 0.89 years; n=89, 58.6% girls; n=125, 82.2% White; n=16, 10.5% Hispanic; median household income US \$90,000-\$100,000) were recruited in 2019 to 2020 from the same area for a randomized controlled trial investigating a mindfulness mobile app. A total of 89.5% (136/152) of the participants completed the trial (for more details, see the study by Hilt et al [48]).

Sample 3 included 88 first-year undergraduate students aged 18 to 21 years (mean age 18.51, SD 0.64 years; n=58, 66% female; n=59, 67% White; n=10, 11% Hispanic; income data

not collected) recruited in 2018 from Lawrence University in Appleton, Wisconsin, for a randomized controlled trial investigating a mindfulness mobile app. They were not oversampled for any characteristic, such as elevated rumination. A total of 90% (79/88) of the participants completed the trial (for more details, see the study by Hilt et al [49]).

Sample 4 included 92 adults from the Madison, Wisconsin, area with elevated levels of depression or anxiety (ie, Patient-Reported Outcomes Measurement Information System [PROMIS] depression or anxiety T-scores of >55 [50,51]) aged 18 to 80 years (mean age 31.28, SD 13.12 years; n=75, 82% female; n=64, 70% White; n=8, 9% Hispanic; median household income US \$45,000, IQR US \$25,000 to US \$125,000) recruited in 2021 for a randomized trial investigating a meditation-based mobile app. In total, 97% (89/92) completed the trial (more details have been provided [52]).

Power analyses for sample 1 [1], sample 2 [48], and sample 3 [49] have been previously reported. Sample 4 (unpublished) was designed to test the feasibility and acceptability of manipulating practice dosage. It was not powered to detect a particular effect. The preregistration can be found on the internet [52,53].

Ethical Considerations

All adult participants completed informed consent procedures approved by the Institutional Review Board at Lawrence University (samples 1-3; IRBs #51418, #72518, and #8616) or the University of Wisconsin-Madison (sample 4; IRB #2019-1578; including agreement that deidentified data from the study could be shared publicly for the latter sample), and the respective institutional review boards approved each study. Given that samples 1 and 2 included minors, parents or guardians completed informed consent procedures, and youth provided assent. Data for all samples were deidentified before data processing and analysis. With regard to compensation, sample 1 received up to US \$90 (US \$30 at baseline and US \$15 at the postintervention time point plus up to US \$15 for app compliance and US \$15 at each of the 2 follow-up periods). Sample 2 received up to US \$105 (US \$30 at baseline and US \$25 at the postintervention time point plus up to US \$15 for app compliance and US \$15 at each of the 2 follow-up periods). Sample 3 received up to US \$30 (US \$5 for each assessment, US \$10 for the intervention period, and up to a US \$5 bonus for app use). For sample 4, participants were paid up to US \$170 for completing all study activities.

Procedure

Adolescent participants (samples 1 and 2) were recruited through letters sent to parents in the local school district and via word of mouth. They were invited to participate if a 2-item phone screen for rumination determined that they reported responding to sadness or stress with rumination at least sometimes. At baseline, adolescents and parents completed self-report questionnaires, set up the mindfulness mobile app (Child and Adolescent Research in Emotion [CARE] app) on the adolescent's mobile device, and learned how to use it. In sample 1, all participants used the same version of the app, which prompted participants to report on mood and rumination and

included a mindfulness intervention. In sample 2, adolescents were randomly assigned to use either a mood monitoring control version of the app or the mindfulness condition of the app that sample 1 used (see the Conditions section). For 3 weeks, adolescents in both samples were prompted to use the CARE app 3 times a day by notifications sent through the app. After the intervention period, participants completed online follow-up questionnaires that included the same measure of trait rumination completed as part of their baseline questionnaires.

College students (sample 3) were recruited from introductory classes, at events for first-year students, and using flyers. Participants were eligible to take part if they did not report serious suicidal concerns. Participants completed baseline questionnaires that included a measure of trait rumination and were randomized to either the mindfulness version of the CARE app or the mood monitoring control condition at baseline. During the 3-week intervention period, participants were prompted to use the app 3 times a day by notifications sent to their phone. At the postintervention time point, participants completed questionnaires that included the same measure of trait rumination completed at baseline.

Adults with elevated depression or anxiety (sample 4) were recruited through flyers placed in the community and recruitment emails sent to faculty, staff, and students at the University of Wisconsin–Madison. Participants were eligible if they were aged ≥ 18 years; had access to a smartphone capable of downloading the Healthy Minds Program (HMP) app; self-reported willingness to complete EMA for 4 weeks; were able to speak, read, and write in English; and had clinically elevated PROMIS depression or anxiety T-scores (>55). Exclusion criteria included previous meditation retreat experience, a regular meditation practice (weekly practice for >1 year or daily practice within the previous 6 months), previous practice under the instruction of a meditation teacher, severe depression symptoms (PROMIS depression T-score of >70), or a positive screen for alcohol dependence on the Alcohol Use Disorders Identification Test [54].

Conditions

Mood Monitoring Control Condition

Samples 2 and 3 used a mood monitoring control condition through the CARE app. Participants assigned to the mood monitoring control condition were prompted to report on state mood and rumination 3 times a day without the chance of receiving a mindfulness intervention. After receiving a notification to use the app, participants reported on their current mood (ie, sad, anxious, happy, and calm) and state rumination on a scale from 0 (*not at all*) to 100 (*extremely*). The work by Hilt et al [48] provides more information.

Meditation Condition

Samples 1, 2, and 3 used a mindfulness meditation condition through the CARE app. Participants assigned to the mindfulness condition answered the same mood and rumination questions as those assigned to the mood monitoring control condition, and they also received mindfulness exercises. To prevent participants from learning which responses would result in receiving a mindfulness exercise, there was a 67% probability

of receiving an exercise each time the app was used. If participants indicated high levels of anxiety or sadness (ie, ≥ 90 on the 0–100 scale), their chances of receiving a mindfulness exercise increased to 85%. If prompted to receive a mindfulness exercise, participants could select how much time they had to complete the exercise (ie, approximately 1, 5, or 10 minutes), and an exercise was randomly assigned within those parameters. After completing the mindfulness exercise, participants reported again on their current mood and rumination. The work by Hilt et al [48] provides more details.

Sample 4 used the 4-week Foundations program from the HMP mobile app. HMP includes meditation practices linked with 4 dimensions of well-being: awareness, connection, insight, and purpose [55]. Briefly, awareness practices aim to cultivate mindfulness and attention regulation, connection practices aim to cultivate healthy relationships with oneself and others, insight practices aim to cultivate an understanding of how our internal experiences (eg, emotions and thoughts) shape our wellbeing, and purpose practices aim to cultivate a connection with one's values and a sense of meaning in daily life. The works by Goldberg et al [56] and Hirshberg et al [57] provide more details. Within this study, participants were randomly assigned to use HMP for either 5 or 15 minutes per day (defined as low- and high-dose conditions, respectively). In addition to using HMP, participants were sent EMAs assessing various aspects of well-being (awareness, connection, insight, and purpose dimensions), psychological distress (depression and anxiety), stressor exposure, and rumination 4 times per day over the 4-week intervention period.

Measures

Self-Report Depressive Symptoms

For adolescents (samples 1 and 2), depressive symptoms were assessed using the Children's Depression Inventory (CDI) [58]. The CDI is a 27-item measure of depression adapted for children and adolescents from the Beck Depression Inventory [12]. The CDI assesses the frequency and severity of depressive symptoms over the previous 2 weeks. Participants report on depressive symptoms and their severity using a 4-point scale, with higher scores on the CDI indicating greater frequency and severity of symptoms. Previous research demonstrates that the CDI is reliable and valid in child and adolescent samples [59,60]. In sample 1, the CDI showed good reliability at baseline (Cronbach $\alpha=0.82$), the postintervention time point (Cronbach $\alpha=0.86$), the 6-week follow-up (Cronbach $\alpha=0.88$), and the 12-week follow-up (Cronbach $\alpha=0.91$). In sample 2, the CDI showed excellent reliability at baseline (Cronbach $\alpha=0.90$), the postintervention time point (Cronbach $\alpha=0.92$), the 6-week follow-up (Cronbach $\alpha=0.91$), and the 12-week follow-up (Cronbach $\alpha=0.92$).

The Beck Depression Inventory–II (BDI-II) [12] was used to assess depressive symptoms in college students (sample 3). The BDI-II is a 21-item measure that asks participants to self-report on depressive symptoms experienced in the previous 2 weeks using a 4-point scale. Higher scores on the BDI-II indicate greater symptom frequency and severity. The BDI-II has demonstrated validity and reliability in college samples [61]. The measure demonstrated excellent reliability at baseline

(Cronbach $\alpha=0.94$), the postintervention time point (Cronbach $\alpha=0.93$), the 6-week follow-up (Cronbach $\alpha=0.88$), and the 12-week follow-up (Cronbach $\alpha=0.91$). For samples 1 to 3 (all of which used the same CARE app), missingness for depression measures was 0.3% (1/320) at baseline and 6.9% (22/320) at the postintervention time point.

The computer-adaptive version of the PROMIS depression scale [62] was used to assess depressive symptoms in adults with elevated depression or anxiety (sample 4). The computer-adaptive PROMIS depression scale draws on a bank of 28 items. Participants rate their experience of various depression symptoms in the previous 7 days on a 5-point scale ranging from 1 (*never*) to 5 (*always*). The measure has shown strong convergent validity with legacy measures of depression, including the BDI-II [50]. As not all participants received the same items, internal consistency could not be calculated. However, short-form versions of the PROMIS depression scale have shown excellent reliability (eg, Cronbach $\alpha=0.97$ for the 8-item PROMIS depression scale [63]). Missingness was very low for sample 4—0% at baseline and 3% (3/92) at the postintervention time point.

Self-Report Trait Rumination

Adolescent self-reported trait rumination (samples 1 and 2) was assessed using the 13-item rumination subscale of the CRSQ [17]. For each item, participants are instructed to report on how they usually respond to feelings of sadness or stress. Instructions were modified to include stress in line with current conceptualizations of rumination [64]. In response to each item, participants indicate whether they respond in the way described by the item on a scale from 0 (*almost never*) to 4 (*almost always*). Higher scores indicate greater rumination. Previous research suggests that the CRSQ is reliable and valid in an adolescent sample [17]. In this study, the rumination subscale of the CRSQ demonstrated good reliability in sample 1 (Cronbach $\alpha=0.89$) and excellent reliability in sample 2 (Cronbach $\alpha=0.92$) at baseline. At the postintervention time point, the CRSQ showed excellent reliability in sample 1 (Cronbach $\alpha=0.90$) and sample 2 (Cronbach $\alpha=0.92$).

College student rumination (sample 3) was assessed using the Ruminative Response Scale (RRS) from the RSQ [65]. The RRS is a 22-item measure in which participants are asked to rate whether they generally respond as described by the item on a scale from 0 (*almost never*) to 4 (*almost always*). Higher scores indicate greater frequency and severity of rumination. Previous research in college student samples suggests that the RRS has good internal consistency and moderate test-retest reliability [66]. Similar to the CRSQ, instructions were adapted to ask participants to report on their response to sadness or stress [64]. In this sample, the RRS showed excellent reliability at baseline (Cronbach $\alpha=0.93$) and follow-up (Cronbach $\alpha=0.94$). For samples 1 to 3, missingness for rumination measures was 0% at baseline and 6.6% (21/320) at the postintervention time point.

Repetitive negative thinking in the adult sample with elevated depression or anxiety (sample 4) was assessed using the Perseverative Thinking Questionnaire (PTQ) [67]. It should be noted that repetitive negative thinking is a broader construct

than rumination (not solely focused on past-oriented repetitive negative thoughts but also includes, eg, future-oriented worry and present concerns). The PTQ is a 15-item measure in which participants are asked how often they respond in a particular way on a scale from 0 (*never*) to 4 (*almost always*). Items query various forms of repetitive negative thinking (eg, “I keep thinking about the same issue all the time”). Similar to the CRSQ and RSQ measures, participants are asked to rate how they *typically* respond. Higher scores indicate greater frequency of repetitive negative thought. Internal consistency and test-retest reliability, along with convergent and predictive validity, have been established for adults [67]. In this sample, the PTQ showed excellent reliability at baseline (Cronbach $\alpha=0.94$) and follow-up (Cronbach $\alpha=0.95$). PTQ missingness for sample 4 was 0% at baseline and 3% (3/92) at the postintervention time point.

EMA Rumination

In samples 1, 2, and 3, state rumination was assessed by asking participants the following questions—“How much were you focusing on your emotions?” and “How much were you focusing on your problems?”—*just before* seeing the prompt to use the app. Participants rated the degree to which they had been ruminating as described by these questions on a scale from 0 (*not at all*) to 100 (*extremely*). These questions were created in line with previous research [68,69]. In the analyses below, we focus on the latter question (problem-focused rumination) to assess rumination given (1) some concerns about the extent to which focusing on emotions in fact assesses problematic rumination (eg, participants may direct *greater* attention to their emotions by virtue of mindfulness training) [7] (see also the study by Nolen-Hoeksema et al [64], which focused on problem-focused rumination); (2) the relatively modest correlation between problem-focused and emotion-focused rumination ($r=0.41$) in this sample, which raises questions regarding internal consistency; and (3) the fact that these 2 items have yielded different patterns of findings in previous work [7]. [Multimedia Appendix 1](#) [70] provides additional analyses based on the emotion-focused item. Participants in samples 1 to 3 had a mean survey compliance of 74% (46.71/63; SD 14.52), 79% (50.02/63; SD 14.24), and 72% (45.45/63; SD 17.78) during the 3-week intervention period, respectively. There was a nonsignificant trend for week-to-week decreases in EMA compliance ($\beta=-0.043$; $P=.10$). [Multimedia Appendix 1](#) provides analyses showing no significant differences in EMA compliance during COVID-19 lockdowns (only 1 sample included participants enrolled during COVID-19 lockdowns).

In sample 4, participants first completed a stressor exposure item that asked them to “Think about the most stressful or negative thing that happened since you completed the last survey” and then indicate how stressed they felt “at the worst point” on a scale from 1 (*not at all*) to 7 (*very much*). Rumination was then assessed using a subsequent item—“After the stressful or negative thing happened, I was dwelling on my mistakes, failures, or losses”—which was also rated on a scale from 1 (*not at all*) to 7 (*very much*). These items were also created in line with previous research [71,72]. EMA compliance was 80.2% (89.85/112; SD 20.6%) and decreased over time by 3.4 percentage points per week ($b=-0.034$; $P<.001$).

Clinical Trial Registration

The data analytic plan for this study was not preregistered. The original trial registration for sample 2 can be found at ClinicalTrials.gov (NCT03900416). The original trial registration for sample 4 can also be found at the Open Science Framework [53] and ClinicalTrials.gov (NCT05229406).

Analytic Strategy

Reliability of Self-Report and EMA Measures

Before estimating the reliability of rumination change scores, we first report the reliability (internal consistency) of conventional retrospective self-reported rumination (for brevity, we refer to the latter simply as “retrospective rumination” in the following sections) at baseline (T1) and the postintervention time point (T2). Next, to compute the reliability of residualized change in retrospective rumination, we used the following formula, which has the reliability of retrospective rumination scores at T1 (α_{T1}) and T2 (α_{T2}) and the squared correlation coefficient ($\rho_{T1,T2}^2$) between these 2 scores [73] as inputs:

$$\text{Reliability}(Z) = \frac{\alpha_{T2} + \rho_{T1,T2}^2 \alpha_{T1} - 2\rho_{T1,T2}^2}{1 - \rho_{T1,T2}^2}$$

It should be noted that the reliability of residualized change (Z) will increase if the reliability of either T1 or T2 rumination scores increases or if the correlation between T1 and T2 rumination *decreases* (see [Multimedia Appendix 1](#) for the reliability of raw rather than residualized change). Next, to approximate the reliability of EMA rumination, we computed a type of split-half reliability estimate evaluating the reliability of the participant mean EMA scores over time [74,75]. Specifically, we *randomly* split each participant’s vector of EMA observations into 2 subvectors, computed mean rumination in each subvector for each participant, and correlated these scores. To approximate reliability of change over time, we computed the ordinary least squares (OLS) linear slope for rumination within each subvector (ie, for each participant, we computed the slope in the 2 random subsets of EMA observations) and correlated these slopes. As each of these correlations is based on quantities derived from only half the available observations, they should be underestimates of reliability following traditional psychometric theory. As a result, a Spearman-Brown (SB) adjustment has been applied in the past [76], although some have questioned the application of this adjustment in this context [75]. Thus, in each instance, we report both the correlations from the random splits both before and following the application of the SB adjustment.

An alternative approach to evaluating the reliability of OLS slopes estimates this reliability *within* a multilevel model (MLM) that includes individual-level random intercepts and slopes (see the work by Raudenbusch and Bryk [77] for details). Consistent with the aforementioned speculation, there is evidence suggesting that reliability estimates are higher for this second approach compared to simply reporting split-half correlations without adjustment [75]. We applied MLM methods to similarly demonstrate this distinction, although we consistently report the former split-half approach both before and following SB

correction to provide greater consistency to the approach taken with self-report measures and because of its frequent use in practice. Another method for evaluating change in this context would attend to individual-level empirical Bayes estimates of the slopes. Such estimates can be viewed as OLS estimates adjusted for lack of reliability. This approach was not taken in this study due to the tendency for such estimates to be “shrunk,” thus potentially introducing bias into analyses evaluating the relationships between slopes and other variables (predictor of either the slopes or outcomes) [78] ([Multimedia Appendix 1](#)).

Given that the measures of reliability for retrospective rumination (internal consistency) and EMA rumination (split-half reliability) remain somewhat different, they are not directly comparable. An important aspect of their distinction is that each attends to different sources of error variability in how the reliability of change is characterized. As noted previously, our self-report reliability estimate attends to item-related error, which is ignored (or irrelevant) in the EMA approach due to the availability of only 1 item; similarly, time-related error is ignored (or irrelevant) in the self-report approach due to a definition of change that incorporates time-related error into true change. In the analyses presented in [Multimedia Appendix 1](#), we demonstrate (using one of our EMA datasets in which a second item could be included) how a data collection design that allows for quantification of both sources of error allows for the assessment of various reliability coefficients that could be applied to estimate reliability coefficients relevant to different conditions. Unfortunately, none of these would be applicable to our evaluation of the reliability of the OLS slope, so they are not included in our main analyses. However, we can evaluate the decrement (if any) in reliability within each measure when examining change over time.

Relationship Between Conventional Self-Report and EMA Measures

Similar to what was mentioned previously, before testing the association between change scores, we first tested the correlation between baseline (T1) scores on the retrospective rumination measure and mean EMA rumination over the initial 3 days of EMA. Similarly, we correlated retrospective rumination at the postintervention time point (T2) and mean EMA rumination over the previous 3 days. We selected a 3-day window for averaging in an effort to obtain a relatively representative estimate of an individual’s typical tendency to ruminate (see [Multimedia Appendix 1](#) for analyses averaging over different numbers of days, which yielded similar findings). We also tested the association between change in retrospective rumination and EMA rumination from T1 to T2. To compute change in retrospective rumination, we saved the residuals from a model in which T2 retrospective rumination scores served as the dependent variable and T1 retrospective rumination was the predictor variable, creating a residualized change score. We focused on residualized change scores because they control for effects of regression to the mean, providing estimates of change that are independent of initial status (in addition to typically having slightly better reliability). Biostatisticians have often argued against using raw change scores for these and other reasons [79]. To compute change in EMA rumination, the slope

of rumination change over the course of the trial (ie, 3 weeks or 21 days for samples 1, 2, and 3 and 4 weeks or 28 days for sample 4) was computed from participant-specific regressions of rumination scores on intervention day [74,75]. It should be noted that, as our focus in this analysis was solely on obtaining and interpreting participant-level estimates of systematic linear change over time, alternative multilevel reliability approaches were perceived as less relevant. However, the section on multilevel reliability in [Multimedia Appendix 1](#) illustrates the application of such methods. Given the differences between samples 1 to 3 (adolescents and college students who all used the identical CARE mindfulness app for 3 weeks) and sample 4 (adults who used the HMP meditation app for 4 weeks), all analyses were conducted separately for the CARE (with sample included as a covariate in the analyses) and HMP studies (henceforth referred to as the CARE and HMP samples, respectively).

Comparing Rumination Improvement From Self-Report and EMA Measures

Given that the CARE sample included a mood monitoring control condition, we tested whether group differences emerged in change in rumination from T1 to T2. For retrospective rumination, we ran a regression predicting T2 rumination (adjusting for T1 rumination scores) with group as the predictor of interest. For EMA rumination, a robust linear mixed-effects model (LMM) using the statistical package *robustlmm* (version 3.1) in R (version 4.2.2; R Foundation for Statistical Computing) with a group $\times$ time interaction was specified. We specified random intercepts and slopes for participants with an unstructured covariance matrix for the random effects, allowing for correlation between random intercepts and slopes within individuals. The residual errors were assumed to have an independent covariance structure (conditional independence given the random effects). These covariance specifications are the defaults in the *robustlmm* package and were used consistently across all robust LMM analyses. A standardized effect size for the group $\times$ time interaction was estimated using the $d_{\text{growth-modeling analysis} - \text{raw}} (\beta_{11}[\text{time}]/\text{SD}_{\text{raw}})$ formula recommended by Feingold [80] (for brevity, $d_{\text{growth-modeling analysis} - \text{raw}}$ is referred to simply as d in the following sections).

For the HMP sample, in which both groups had access the HMP app, we examined whether EMA rumination changed over time. To do this, we fit a robust LMM with EMA rumination as the dependent variable and time in days (ranging from 0 to 28) as the independent variable, with the nesting of observations within participants modeled using a random intercept and slope. A standardized effect size for the effect of time was estimated following the aforementioned $d_{\text{growth-modeling analysis} - \text{raw}}$ formula.

Criterion Validity of Self-Report and EMA Measures

Finally, to test whether change in traditional self-reported rumination or EMA rumination predicted improvement in depression severity (criterion), we used robust LMMs. Consistent with prior work arguing that targeting repetitive negative thinking (rumination) improves depressive symptoms [6,7,81], these analyses were an attempt to test whether improvement in either retrospective rumination or EMA

rumination were related to improvement in depression. For the CARE sample, the dependent variable incorporated repeated assessment of depression over 4 time points (ie, baseline, posttreatment time point, 6-week follow-up, and 12-week follow-up) that were nested within individuals. Time was centered to represent estimated symptom scores at the final follow-up (12 weeks), and baseline depression was included as a covariate. Random intercepts and slopes were specified. To test whether either change measure (baseline to posttreatment time point) predicted improvement in depression over time (baseline to the 12-week follow-up), we included each term in an interaction with time (ie, retrospective rumination change $\times$ time and EMA rumination change $\times$ time). Each term was simultaneously included in the same model.

For the HMP sample, traditional self-reported depression was assessed only at the pre- and posttest time points. Therefore, we used OLS regression models with posttest depression as the dependent variable and pretest depression, retrospective rumination change, and EMA rumination change entered as predictors.

To reduce the influence of outliers on both the CARE and HMP samples, we winsorized any extreme values (*winsorize* function in the *DescTools* R package). Correlation and regression analyses used listwise deletion for the specific variables included in each model. Missing EMA observations within participants were handled using maximum likelihood estimation in mixed-effects models, which uses all available data and provides unbiased estimates under the missing at random assumption. Data were analyzed using R (version 4.2.2).

For correlation analyses between conventional self-report and EMA measures, our combined CARE sample ($n=320$) provided $>80\%$ power to detect correlations of $r \geq 0.16$ at an α value of .05. For the HMP sample ($n=92$), we had $>80\%$ power to detect correlations of $r \geq 0.29$ (using the *pwr* package in R). For analyses testing group differences in conventional self-reported rumination improvement in the CARE sample, we had $>80\%$ power to detect a small effect size (Cohen $f^2 \geq 0.025$) for between-group comparisons (*pwr* package). For mixed-effects models testing group differences in EMA rumination trajectories over time (condition $\times$ day interaction) in the CARE sample, we had $>80\%$ power to detect a small effect size of Cohen $d \geq 0.25$ (using the *mixedpower* package). For the HMP sample, all participants had access to the HMP app (no control group), so we tested within-subject rumination change over time. We had $>80\%$ power to detect a small to medium effect size of Cohen $d \geq 0.30$ for paired t tests of pretest-posttest rumination change (*pwr* package) and Cohen $d \geq 0.22$ (small effect size) for mixed-effects models testing EMA rumination decline over the study period (*mixedpower* package).

Data Transparency Statement

Data were drawn from 4 clinical trials. Primary outcome data have been published for 3 of these trials. Primary results from the fourth trial have not yet been published. This study aggregated data across these 4 trials to evaluate the psychometric properties of EMA versus conventional self-report measures

for improvement in rumination. This question has not yet been addressed within these 4 studies.

Results

Overview

Table 1 shows the participant demographic characteristics.

Table 1. Demographic characteristics of participants at baseline.

Characteristic	Sample 1 (n=80), n (%)	Sample 2 (n=152), n (%)	Sample 3 (n=88), n (%)	Sample 4 (n=92), n (%)
Sex				
Male	43 (53.8)	63 (41.4)	26 (29.5)	13 (14.1)
Female	36 (45)	89 (58.6)	58 (65.9)	75 (81.5)
Nonbinary	0 (0)	0 (0)	4 (4.5)	4 (4.3)
Chose not to answer	1 (1.3)	0 (0)	0 (0)	0 (0)
Race				
American Indian or Alaska Native	2 (2.5)	0 (0)	0 (0)	2 (2.2)
Asian	0 (0)	3 (2)	18 (20.5)	18 (19.6)
Black or African American	1 (1.3)	5 (3.3)	8 (9.1)	3 (3.3)
Multiracial	1 (1.3)	16 (10.5)	0 (0)	4 (4.3)
Native Hawaiian or Pacific Islander	0 (0)	1 (0.7)	0 (0)	0 (0)
White	69 (86.3)	125 (82.2)	59 (67)	64 (69.6)
Chose not to answer	7 (8.8)	2 (1.3)	3 (3.4)	1 (1.1)
Ethnicity				
Hispanic	3 (3.8)	16 (10.5)	10 (11.4)	8 (8.7)
Non-Hispanic	75 (93.8)	136 (89.5)	78 (88.6)	83 (90.2)
Chose not to answer	2 (2.5)	0 (0)	0 (0)	1 (1.1)

Reliability of Self-Report and EMA Measures

CARE Samples

As internal consistency is computed based on the raw item scores, the 2 adolescent samples completed the CRSQ, and the other (college student) sample used the RSQ, we computed reliability separately for both measures. Reliability (internal consistency) for retrospective rumination was high at T1 (Cronbach α =0.89 for adolescents and Cronbach α =0.93 for college students) and T2 (Cronbach α =0.90 for adolescents and Cronbach α =0.94 for college students) but decreased for the residualized change score (r =0.71 for adolescents and 0.84 for college students). Reliability (split-half) for EMA rumination was high (r =0.89; SB adjusted=0.94) but decreased substantially for the change score (r =0.50; SB adjusted=0.67; using the reliability estimate from the MLM: r =0.58). There were fewer EMA time points per participant in the CARE samples (mean 50.2, SD 17.6; median 55.0) relative to the HMP sample (mean 89.9, SD 23.1; median 98). Reliability estimates are often higher when there are more time points as the slopes are estimated with less error. We recomputed reliability for the change score excluding individuals with <40 time points, which increased the reliability coefficients from r =0.50 (SB adjusted=0.67) to r =0.61 (SB adjusted=0.76).

HMP Sample

Reliability (internal consistency) for retrospective rumination was high at T1 (Cronbach α =0.94) and T2 (Cronbach α =0.95). The residualized change score in this sample also retained high reliability (r =0.90). Reliability (split-half) for EMA rumination was high (r =0.96; SB adjusted=0.98) but decreased for the change score (r =0.77; SB adjusted=0.87; it was also 0.87 using the MLM-based reliability estimate). It should be noted that the relatively high residualized change reliabilities for the HMP and CARE college student samples were due to the high reliabilities of the constituent T1 and T2 rumination scores (see the previous paragraph) and their relatively low intercorrelation (CARE college student r =0.66; HMP r =0.55 vs CARE adolescent r =0.69), implying more substantial variability in real change (see, eg, formula 7.10 in the work by Crocker and Algina [82]). Multimedia Appendix 1 provides additional analyses, including convergent and divergent validity tests for EMA and the reliability of raw rather than residualized change scores that were similar.

Correlation Between Self-Report and EMA Measures

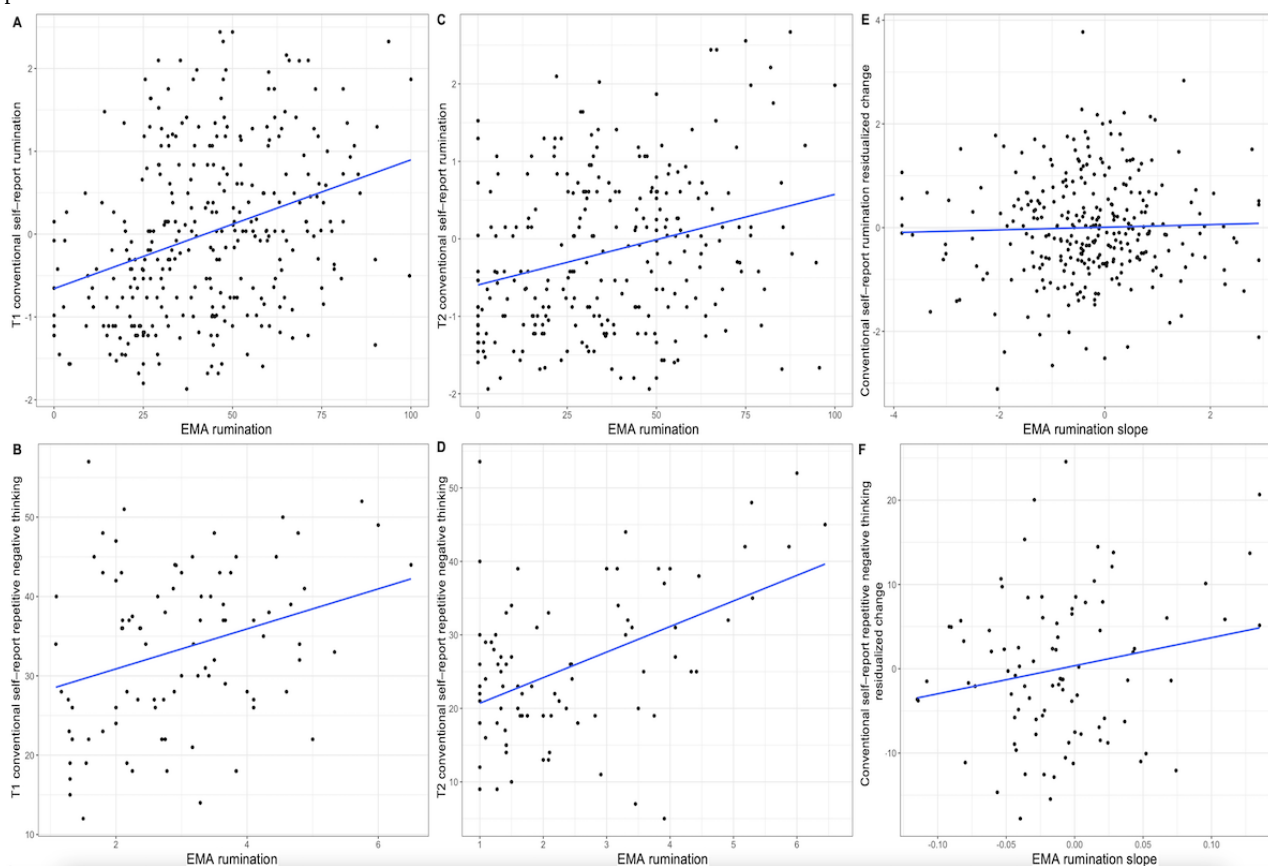
CARE Samples

Retrospective rumination at T1 was significantly positively correlated with mean EMA rumination over the subsequent 3 days (r =0.35; P <.001; Figure 1A). Similarly, the correlation between retrospective rumination at T2 and EMA rumination

over the previous 3 days was significant ($r=0.28$; $P<.001$; Figure 1C). The magnitude of these correlations is conventionally considered to be in the medium range [83].

Residualized T1 to T2 change in retrospective rumination was not significantly correlated with change (slope) in EMA rumination ($r=0.03$; $P=.66$; Figure 1E).

Figure 1. Scatterplot of the association between (1) conventional retrospective rumination at T1 and ecological momentary assessment (EMA) rumination over the subsequent 3 days for the Child and Adolescent Research in Emotion (CARE; panel A) and Healthy Minds Program (HMP; panel B) samples, (2) conventional retrospective rumination at T2 and EMA rumination over the previous 3 days for the CARE (panel C) and HMP (panel D) samples, and (3) residualized T1 to T2 change in retrospective rumination and change (slope) in EMA rumination for the CARE (panel E) and HMP (panel F) samples.



HMP Sample

Retrospective rumination at T1 was significantly positively correlated with mean EMA rumination over the first 3 days of EMA ($r=0.31$; $P=.003$; Figure 1B). Similarly, the correlation between retrospective rumination at T2 and EMA rumination over the final 3 days of EMA was significant ($r=0.47$; $P<.001$; Figure 1D). These correlations were medium to large in magnitude [83].

Residualized T1 to T2 change in retrospective repetitive negative thinking was not significantly correlated with change (slope) in EMA rumination ($r=0.20$; $P=.06$; Figure 1F).

Comparing Rumination Improvement From Self-Report and EMA Measures

CARE Samples

A significant group difference emerged for change in retrospective rumination such that the mindfulness group exhibited significantly greater improvement relative to the mood monitoring control group ($d=0.37$; $b=-0.34$; SE 0.096; $t_{294}=-3.51$; $P<.001$). In contrast, there were no group differences in EMA rumination change ($d=0.14$; $b=-0.20$; SE 0.138; $P=.14$).

Overall change in the CARE trials (averaging across both conditions) was small for retrospective rumination ($d=0.19$) and EMA rumination ($d=0.20$).

HMP Sample

There was a significant pretest-posttest reduction in retrospective repetitive negative thinking in the HMP sample ($d=0.77$; $t_{88}=7.60$; $P<.001$). EMA rumination decreased over time ($b=-0.012$; SE 0.0050; $P=.02$). However, the associated effect size was small ($d=0.17$).

Predictive Validity of Self-Report and EMA Measures

CARE Samples

Greater pre- to postintervention improvement in both retrospective rumination ($b=0.04$; SE 0.017; $P=.03$) and EMA rumination ($b=0.03$; SE 0.014; $P=.04$) were significantly positively associated with improvement in depressive symptoms from baseline to follow-up.

HMP Sample

Similarly, greater pre- to postintervention improvement in both retrospective rumination ($b=0.23$; SE 0.060; $P<.001$) and EMA rumination ($b=0.2049$; SE 9.94; $P=.04$) were significantly

positively associated with improvement in depression symptoms from baseline to the posttest time point.

Discussion

Principal Findings

Overview

Researchers are increasingly incorporating EMA into intervention studies in an effort to obtain a more fine-grained and ecologically valid assessment of change. EMA can provide high-resolution information about treatment-relevant processes (eg, emotional, cognitive, and behavior change) while minimizing memory biases relative to conventional (retrospective) self-report measures. However, repeated EMA surveys can be burdensome to participants, which could lead to poor compliance; careless responding (eg, random or stereotyped responses to items); or a bias toward the enrollment of participants who are more motivated, thus limiting generalizability. In addition, the psychometric properties of EMA measures are often not well established before implementing them in a study (eg, researchers often create new EMA items to assess a construct of interest or adapt them from an existing self-report scale without evaluating or reporting their psychometric properties) [84]. In short, intervention researchers must carefully consider whether the benefits of implementing an EMA measure in a study outweigh potential costs and limitations. In this study, we attempted to help address the question of whether EMA is worth including by combining data from 4 trials (total $N=412$), all of which examined a meditation app and measured outcomes (change in rumination) via both conventional retrospective self-report questionnaires and EMA. We did so by evaluating 4 relevant criteria.

Criterion 1: Are EMA Measures of Rumination Reliable?

While the reliability of both retrospective rumination (at baseline and the postintervention time point) and mean EMA rumination scores was high, our findings revealed a decline in reliability when assessing change over time, especially for EMA. These results align with long-standing concerns about the unreliability of measures being compounded when measuring change [85,86]. Reliability is generally defined as the ratio of systematic (“true”) variance of interest to total observed variance. In addition to high measurement error, another reason for the low reliability in change scores is that true change variability may be low, either because of limited change or because all participants change to a nearly equivalent degree [87]. Previous EMA studies have reported relatively lower reliability for change scores [88–90]. Critically, low reliability can result in biased inferences and a loss of statistical power, leading to an inaccurate estimation of the true effect on the population [91]. Consequently, when interpreting findings, it becomes extremely challenging to determine whether the results are trustworthy or whether they are attributable to measurement error, thereby reducing the likelihood of successfully replicating findings [92]. Despite the importance of establishing reliability, many EMA studies do not adequately report the psychometrics of their measures [84].

In theory, 2 approaches to improving the reliability (and likely the validity) of EMA measures are to increase the number of time points or add more items that capture the construct of interest [93]. In this study, we relied on single-item measures of EMA rumination, which may provide less coverage of the construct of interest relative to multi-item measures [89]. However, it is important to note that increasing the number of items or assessment time points also increases the burden on participants and, consequently, may reduce compliance or increase careless responding [94]. Therefore, researchers should balance the need to improve measurement (via more items or assessment time points) with a careful consideration of participant burden. In this study, using single-item measures (a common practice in many EMA studies) also prevented us from calculating internal consistency, the method we used to compute reliability for the conventional self-report measures. Consequently, we were unable to directly compare the reliabilities of the conventional self-report and EMA measures.

Similar to increasing the number of relevant items, averaging scores across multiple measurement time points is known to improve reliability [84,93]. This approach aligns with our finding of high reliability for *mean* EMA rumination. For example, in the context of studying day-to-day changes in rumination, averaging data collected from multiple within-day occasions (as opposed to relying on a single measurement per day) to estimate rumination levels each day is expected to improve the precision of within-day rumination. In addition, the higher reliability of EMA change scores in the HMP sample relative to the CARE sample may be due to the larger number of time points collected per participant in the former sample, improving the precision of slope estimates.

In summary, the relatively low reliability of change scores (in particular for EMA measures of rumination) is concerning. It would be helpful to better understand the scope of this issue if researchers more consistently reported the reliability of change where it is evaluated given that this is not common practice yet can have a profound impact on study findings (eg, low reliability attenuating effect sizes). This can aid readers in evaluating EMA findings in a given study in light of potential psychometric limitations. The field may also benefit from additional innovative approaches to assessing the reliability of EMA measures [89].

Criterion 2: Correlation With Traditional Self-Report Measures

The results revealed statistically significant medium to large ($r=0.28$ to 0.47) correlations [83] between rumination assessed via conventional, retrospective self-report and the mean of EMAs. However, there were no significant correlations between changes in retrospective rumination and change in EMA rumination ($r=0.03$ to 0.20). These results are in line with those of previous studies that have reported low to moderate correlations between trait and state measures of various constructs, including anhedonia [95], affective lability [96,97], and personality traits [98,99].

Several factors may account for the relatively modest association between conventional retrospective self-report and EMA measures of rumination (especially for the change scores). First,

less than perfect reliability can attenuate associations between any 2 variables. Reliability estimates were particularly low for change scores, which may explain (at least in part) the especially low correlation between those measures. Second, the distinction by Kahneman [100] between the “experiencing self” and the “remembering self” provides a useful framework for interpreting these results. According to this distinction, when participants are asked to recall their past experiences on conventional, retrospective self-report measures (in this study, asking about past ruminative thoughts), they may be especially influenced by the most intense (peak) and recent (end) levels of rumination (ie, the so-called peak-end bias) rather than equally weighting and simply averaging all time points, as reflected by the mean of repeated EMAs [27-30]. Third, EMA requires introspection of *current* states, whereas retrospective self-report measures require participants to access autobiographical (episodic) memory and provide an average report of past experiences. The accuracy of these reports is limited by the extent to which past experiences have been accurately encoded and consolidated in memory [11,31,32]. Moreover, when recall time frames are longer, individuals may shift from episodic memory to semantic memory, which includes more abstract beliefs or generalization about the self [11,34]. Thus, when asked to report on their typical tendency to ruminate on self-report measures, individuals’ responses may be influenced by relevant self-concepts (“I tend to overthink” or “I’m a worrier”), how they believe others view them, or how they would like to be perceived [101]. In summary, in addition to reduced reliability, the relatively modest correlations between conventional retrospective self-report and EMA measures of rumination may be due to the fact that the 2 measurement approaches ultimately tap into different conscious or functional “selves” [31]. As described in the following sections, the decision to include EMA or retrospective measures of change in an intervention study should be informed by which of these “selves” (experiencing vs remembering self) the researchers want to assess and expect to change.

Criterion 3: Sensitivity to Change

Compared to conventional retrospective self-report, EMA measures detected smaller group (intervention vs control) differences in rumination improvement (effect sizes: $d=0.37$ vs $d=0.14$ for self-report and EMA, respectively). EMA also showed less linear change in rumination over time in the intervention group (effect sizes: $d=0.77$ vs $d=0.17$ for self-report and EMA in the HMP sample, respectively). These findings have important implications for the selection of outcome measures in intervention studies. They imply that a trial may yield very different findings (eg, significant vs nonsignificant differences in outcomes between the intervention and control group) based on whether a conventional retrospective or EMA instrument was used as the outcome measure. This raises the critical question of which result is closer to the “ground truth.” The larger effect size observed for retrospective self-report could be due to this measure being a more sensitive method for capturing changes in rumination. Although speculative, EMA may be less sensitive to detecting change due the repeated nature of EMA surveys introducing unique forms of bias. For example, perhaps participants tend to align their responses with their

(relatively recent) previous answers (a type of anchoring bias), potentially resulting in reduced detection of actual change. If a participant rated their rumination as a 4 on a scale from 1 to 5 on the last EMA survey, an anchoring bias may lead them to report a similar score on the next survey even if their “true” rumination score is now lower. Alternatively, EMA may provide a more *accurate* estimation of real change (which may, in fact, be modest), whereas the conventional retrospective self-report measure may overinflate estimates of change due to influences such as social desirability bias, regression to the mean, or initial elevation bias [39].

If, as discussed previously, EMA and retrospective self-reports indeed capture related but somewhat different constructs, it is possible that the pattern of results we observed is due to the intervention differentially affecting each of these facets. For example, it may be that the intervention had a larger impact on one’s episodic or semantic representations of being a ruminator (as measured via retrospective self-report) but had a more modest impact on actual moment-to-moment rumination (as measured via EMA). There may also be a causal relationship from the latter to the former. Namely, relatively modest improvements in day-to-day momentary rumination may have a proportionally large positive impact on one’s self-concept of being a ruminator.

Even if changes in retrospective self-report are biased due to the limitations of memory and there is little change in actual experience, these changes in retrospective recall may still be meaningful. Research has shown that future behaviors (eg, whether to return for a colonoscopy screening or end a romantic relationship) are *better* predicted by retrospective measures of experience than by actual momentary experiences [102]. In summary, retrospective self-reports (eg, related to one’s self-concept and tendency to engage in repetitive negative thought) may, of course, be very personally meaningful to the individual who underwent the intervention. Therefore, if the goal is to understand what individuals experienced in the moment, EMA arguably provides a more accurate measure. However, if the focus is on individuals’ global impressions of their experiences and themselves, then retrospective self-report measures may be more valuable.

Criterion 4: Predicting Change in Depressive Symptoms

Finally, the results of this study revealed that both retrospective and EMA measures of rumination change predicted improvement in depressive symptoms, with each measure contributing significantly beyond the effects of the other (ie, evidence of the incremental predictive validity of each measure). These findings suggest that these 2 methods assess meaningfully distinct aspects of rumination, both of which hold clinically important information. Therefore, if feasible, combining retrospective and EMA data could provide a more comprehensive picture of clinically meaningful improvement over time. It is also important to highlight that there are, of course, other reasons why a researcher may want to include EMA measures in an intervention study. For example, EMA enables the examination of symptom or emotion dynamics over time, such as changes in inertia or instability, which cannot be captured using traditional self-report methods [103]. In addition,

EMA enables more sophisticated analyses of mechanisms of change, such as temporal mediation analyses that can examine how changes in putative mediators precede and predict subsequent changes in outcomes within individuals over time, addressing key limitations of traditional mediation approaches [46,104]. Nevertheless, these methodological advantages require careful attention to data quality as poor reliability in EMA can obscure the ability to detect meaningful temporal dynamics in these processes.

Strengths and Limitations

There are several notable strengths to this study. First, we combined data across 4 trials to increase sample size (total $N=412$). Second, each of the trials examined a meditation app and measured outcomes via both retrospective self-report and EMA. Finally, we evaluated several metrics of reliability, validity, sensitivity to change, and incremental predictive validity to examine the value of EMA measures in an intervention study. At the same time, this study had several limitations. First, with regard to criterion 4, a well-validated clinical interview of depressive symptoms would have been preferable to a self-report instrument to serve as the “ground truth” of change during the intervention. Second, in contrast to the CARE samples, which had follow-up time points, the HMP sample analyses for criterion 4 (predicting change in depressive symptoms) were limited in testing the relationship between change in rumination and depressive symptoms over the same time frame (ie, pre- to posttreatment time point), which is problematic [105] relative to lagged analyses testing whether early change in rumination prospectively predicts subsequent depressive symptom improvement. Third, we only focused on rumination. The extent to which our findings generalize to other common outcome measures (eg, comparing conventional self-report and EMA measures of change in depression or anxiety symptoms) is unknown. Fourth, the fact that the items in the retrospective and EMA measures of rumination were not identical likely attenuated correlations. Fifth, it would have been preferable to have a multi-item rather than a single-item measure of EMA rumination, which would have allowed us to compute internal consistency and multilevel reliability estimates [89]. On the other hand, given the burden of EMA, single-item measures are very common, and thus, our study may generalize to the existing literature. Sixth, it may be that mindfulness training (or perhaps even repeatedly answering rumination questions in the control group) shifted how participants interpreted and responded to the rumination prompts [106-108]. For example, on the topic of the influence of meditation on item interpretation and response patterns, there is evidence of differential item functioning among meditators and nonmeditators on mindfulness measures [108]. Thus, it may be that meditation practice (eg, cultivating nonjudgmental awareness of internal experience) shifts the interpretation of mindfulness items, which could also apply to the rating of negative mental states assessed via rumination items (eg, see the previous sections for our concern about the item focusing on emotions). However, the use of only 1 item for measurement in EMA, combined with the relatively low sample sizes, precluded our ability to meaningfully evaluate longitudinal measurement invariance. Seventh, the fact that the included

samples and trials differed in patient characteristics (eg, age) and methods (eg, different measures of rumination) could be considered a limitation. On the other hand, this allowed us to test the consistency of the findings (eg, reliability of and correlations between retrospective and EMA measures) across these samples and study differences (which, overall, were quite consistent, which provides some evidence, albeit limited, for the generalizability of the findings). Finally, although our samples were diverse in age and setting, they were predominantly White (59/88, 67%-69/80, 86%). This limits the generalizability of our findings across racial and cultural groups. Previous research suggests that the strength of association between rumination and psychological outcomes may differ across cultural backgrounds [109].

Conclusions

Despite these limitations, the results of this study provide insights into the utility of conventional self-report and EMA measures of rumination that are relevant considerations for researchers when designing studies and interpreting results. First, the reliability of change in rumination over the course of the intervention was relatively low, especially for EMA. This is concerning given that change over time is precisely what intervention researchers are most interested in. Increasing the number of EMA assessment time points and using multi-item scales while being mindful of not overburdening participants may improve the precision with which change over time is estimated. The frequently observed low reliability of change is often presented as a concern about the measurement of change more generally [110]. However, beyond the reliability of measures at each time point, the most critical factor is the correlation between measures across time points, which can be viewed as an index of the variability of change. As this will be a feature of both study and population conditions that will not be consistent across studies or under the direct control of the investigator, we concur with previous commentaries on the issue that reliability of change scores should be evaluated on a case-by-case basis rather than dismissed out of hand [111-113].

Second, conventional self-report and EMA measures of rumination were modestly correlated at a given time point, with no significant correlation between changes in retrospective and EMA measures of rumination over time. Despite this, changes in both measures of rumination predicted decreased depressive symptoms. This suggests that conventional self-report and EMA measures of rumination are not redundant. Rather, each measure captures distinct and clinically meaningful facets of rumination (ie, remembered vs experienced rumination). Using both conventional self-report and EMA measures of rumination in combination, when feasible, may provide a more comprehensive picture of clinical change. It can be appreciated that the tendency for conventional self-report measures to administer many items at few time points and the tendency for EMA to administer fewer items at many time points also provide complementary designs that minimize the effects of different sources of error (in the case of self-report, item-related error and, in the case of EMA, time-related error), which may be useful to quantify using generalizability theory techniques, and that can be incorporated into a multilevel reliability analysis. There is value for future research to carefully consider such complexities related to the

psychometric properties of EMA measures of rumination and, EMA literature. ideally, ultimately standardize measures of rumination in the

Acknowledgments

This research was supported by National Center for Complementary and Integrative Health grant K23AT010879 (SBG), the American Psychological Foundation (LMH), and National Institute of Mental Health grant R15MH116303 (LMH). The first author (CAW) was partially supported by the National Institute of Mental Health (R01MH116969), the National Center for Complementary and Integrative Health (R01AT011002), the Tommy Fuss Fund, and a National Alliance for Research on Schizophrenia and Depression Young Investigator Grant from the Brain & Behavior Research Foundation. Artificial intelligence was not used in any portion of manuscript writing. During the resubmission, we did use Anthropic's Claude to assist with statistical analysis code in R (eg, how to appropriately specify a random slope and intercept in robust linear mixed-effects models). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or other funders.

Data Availability

The code and datasets generated or analyzed for sample 4 are available in the Open Science Framework repository [114]. The datasets generated or analyzed for samples 1 to 3 are not publicly available as the institutions require a signed data sharing agreement, but are available from the corresponding author on request.

Authors' Contributions

CAW and SBG conceptualized the study and contributed to formal analysis, with CAW additionally handling visualization and SBG contributing to funding acquisition. LMH secured funding, conducted the investigation, and managed data curation alongside CMS, who also handled project administration. DMB contributed to the conceptualization and formal analysis of the data. HF provided input on the study conceptualization. All authors contributed to writing the original draft (CAW writing the bulk of the original draft) and participated in the review and editing process.

Conflicts of Interest

CAW has received consulting fees from the King & Spalding law firm for work unrelated to this publication.

Multimedia Appendix 1

Supplementary results and tables presenting additional analyses.

[DOCX File, 45 KB-Multimedia Appendix 1]

References

1. Hilt LM, Swords CM. Acceptability and preliminary effects of a mindfulness mobile application for ruminative adolescents. *Behav Ther*. Dec 2021;52(6):1339-1350. [doi: [10.1016/j.beth.2021.03.004](https://doi.org/10.1016/j.beth.2021.03.004)] [Medline: [34656190](https://pubmed.ncbi.nlm.nih.gov/34656190/)]
2. Mofsen AM, Rodebaugh TL, Nicol GE, Depp CA, Miller JP, Lenze EJ. When all else fails, listen to the patient: a viewpoint on the use of ecological momentary assessment in clinical trials. *JMIR Ment Health*. May 21, 2019;6(5):e11845. [FREE Full text] [doi: [10.2196/11845](https://doi.org/10.2196/11845)] [Medline: [31066701](https://pubmed.ncbi.nlm.nih.gov/31066701/)]
3. Moore RC, Depp CA, Wetherell JL, Lenze EJ. Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. *J Psychiatr Res*. May 2016;75:116-123. [FREE Full text] [doi: [10.1016/j.jpsychires.2016.01.011](https://doi.org/10.1016/j.jpsychires.2016.01.011)] [Medline: [26851494](https://pubmed.ncbi.nlm.nih.gov/26851494/)]
4. Peterson CB, Engel SG, Crosby RD, Strauman T, Smith TL, Klein M, et al. Comparing integrative cognitive-affective therapy and guided self-help cognitive-behavioral therapy to treat binge-eating disorder using standard and naturalistic momentary outcome measures: a randomized controlled trial. *Int J Eat Disord*. Oct 25, 2020;53(9):1418-1427. [doi: [10.1002/eat.23324](https://doi.org/10.1002/eat.23324)] [Medline: [32583478](https://pubmed.ncbi.nlm.nih.gov/32583478/)]
5. Targum SD, Sauder C, Evans M, Saber JN, Harvey PD. Ecological momentary assessment as a measurement tool in depression trials. *J Psychiatr Res*. May 2021;136:256-264. [FREE Full text] [doi: [10.1016/j.jpsychires.2021.02.012](https://doi.org/10.1016/j.jpsychires.2021.02.012)] [Medline: [33621911](https://pubmed.ncbi.nlm.nih.gov/33621911/)]
6. Webb CA, Swords CM, Murray L, Hilt LM. App-based mindfulness training for adolescent rumination: predictors of immediate and cumulative benefit. *Mindfulness* (N Y). Oct 09, 2021;12(10):2498-2509. [FREE Full text] [doi: [10.1007/s12671-021-01719-0](https://doi.org/10.1007/s12671-021-01719-0)] [Medline: [35432625](https://pubmed.ncbi.nlm.nih.gov/35432625/)]
7. Webb CA, Swords CM, Lawrence HR, Hilt LM. Which adolescents are well-suited to app-based mindfulness training? A randomized clinical trial and data-driven approach for personalized recommendations. *J Consult Clin Psychol*. Oct 2022;90(9):655-669. [FREE Full text] [doi: [10.1037/ccp0000763](https://doi.org/10.1037/ccp0000763)] [Medline: [36279218](https://pubmed.ncbi.nlm.nih.gov/36279218/)]

8. Webb CA, Murray L, Tierney AO, Gates K. Dynamic processes in behavioral activation therapy for anhedonic adolescents: modeling common and patient-specific relations. *J Consult Clin Psychol*. Aug 2024;92(8):454-465. [doi: [10.1037/ccp0000830](https://doi.org/10.1037/ccp0000830)] [Medline: [37276084](https://pubmed.ncbi.nlm.nih.gov/37276084/)]
9. Wichers MC, Barge-Schaapveld DQ, Nicolson NA, Peeters F, de Vries M, Mengelers R, et al. Reduced stress-sensitivity or increased reward experience: the psychological mechanism of response to antidepressant medication. *Neuropsychopharmacology*. Mar 21, 2009;34(4):923-931. [doi: [10.1038/npp.2008.66](https://doi.org/10.1038/npp.2008.66)] [Medline: [18496519](https://pubmed.ncbi.nlm.nih.gov/18496519/)]
10. Stone AA, Shiffman SS. Ecological validity for patient reported outcomes. In: Steptoe A, editor. *Handbook of Behavioral Medicine: Methods and Applications*. New York, NY. Springer; 2010:99-112.
11. Van den Bergh O, Walentynowicz M. Accuracy and bias in retrospective symptom reporting. *Curr Opin Psychiatry*. Oct 2016;29(5):302-308. [doi: [10.1097/YCO.0000000000000267](https://doi.org/10.1097/YCO.0000000000000267)] [Medline: [27427854](https://pubmed.ncbi.nlm.nih.gov/27427854/)]
12. Beck AT, Steer RA, Brown GK. *Manual for the Beck Depression Inventory-II*. San Antonio, TX. Psychological Corporation; 1996.
13. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *Am J Prev Med*. 1994;10(2):77-84. [Medline: [8037935](https://pubmed.ncbi.nlm.nih.gov/8037935/)]
14. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. Mar 01, 1960;23(1):56-62. [FREE Full text] [doi: [10.1136/jnnp.23.1.56](https://doi.org/10.1136/jnnp.23.1.56)] [Medline: [14399272](https://pubmed.ncbi.nlm.nih.gov/14399272/)]
15. Poznanski EO, Mokros HB. *Children's Depression Rating Scale, revised (CDRS-R)*. Western Psychological Services. 1996. URL: <https://eprovide.mapi-trust.org/instruments/children-s-depression-rating-scale-revised> [accessed 2024-05-29]
16. Nolen-Hoeksema S. Responses to depression and their effects on the duration of depressive episodes. *J Abnorm Psychol*. Dec 1991;100(4):569-582. [doi: [10.1037//0021-843x.100.4.569](https://doi.org/10.1037//0021-843x.100.4.569)] [Medline: [1757671](https://pubmed.ncbi.nlm.nih.gov/1757671/)]
17. Abela JR, Brozina K, Haigh EP. An examination of the response styles theory of depression in third- and seventh-grade children: a short-term longitudinal study. *J Abnorm Child Psychol*. Oct 2002;30(5):515-527. [doi: [10.1023/a:1019873015594](https://doi.org/10.1023/a:1019873015594)] [Medline: [12403154](https://pubmed.ncbi.nlm.nih.gov/12403154/)]
18. Chen E, Zeltzer LK, Craske MG, Katz ER. Children's memories for painful cancer treatment procedures: implications for distress. *Child Dev*. Jan 28, 2000;71(4):933-947. [doi: [10.1111/1467-8624.00200](https://doi.org/10.1111/1467-8624.00200)] [Medline: [11016557](https://pubmed.ncbi.nlm.nih.gov/11016557/)]
19. Noel M, Rabbitts JA, Tai GG, Palermo TM. Remembering pain after surgery: a longitudinal examination of the role of pain catastrophizing in children's and parents' recall. *Pain*. May 2015;156(5):800-808. [FREE Full text] [doi: [10.1097/j.pain.0000000000000102](https://doi.org/10.1097/j.pain.0000000000000102)] [Medline: [25630028](https://pubmed.ncbi.nlm.nih.gov/25630028/)]
20. Broderick JE, Schwartz JE, Vikingstad G, Pribbernow M, Grossman S, Stone AA. The accuracy of pain and fatigue items across different reporting periods. *Pain*. Oct 30, 2008;139(1):146-157. [FREE Full text] [doi: [10.1016/j.pain.2008.03.024](https://doi.org/10.1016/j.pain.2008.03.024)] [Medline: [18455312](https://pubmed.ncbi.nlm.nih.gov/18455312/)]
21. Friedberg F, Sohl SJ. Memory for fatigue in chronic fatigue syndrome: the relation between weekly recall and momentary ratings. *Int J Behav Med*. 2008;15(1):29-33. [doi: [10.1007/BF03003071](https://doi.org/10.1007/BF03003071)] [Medline: [18444018](https://pubmed.ncbi.nlm.nih.gov/18444018/)]
22. Giske L, Sandvik L, Røe C. Comparison of daily and weekly retrospectively reported pain intensity in patients with localized and generalized musculoskeletal pain. *Eur J Pain*. Oct 16, 2010;14(9):959-965. [doi: [10.1016/j.ejpain.2010.02.011](https://doi.org/10.1016/j.ejpain.2010.02.011)] [Medline: [20363653](https://pubmed.ncbi.nlm.nih.gov/20363653/)]
23. Rinner MT, Meyer AH, Mikoteit T, Hoyer J, Imboden C, Hatzinger M, et al. General or specific? The memory-experience gap for individuals diagnosed with a major depressive disorder or a social phobia diagnosis, and individuals without such diagnoses. *Memory*. Oct 16, 2019;27(9):1194-1203. [doi: [10.1080/09658211.2019.1640252](https://doi.org/10.1080/09658211.2019.1640252)] [Medline: [31311430](https://pubmed.ncbi.nlm.nih.gov/31311430/)]
24. Stone AA, Broderick JE, Shiffman SS, Schwartz JE. Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain*. Jan 2004;107(1-2):61-69. [doi: [10.1016/j.pain.2003.09.020](https://doi.org/10.1016/j.pain.2003.09.020)] [Medline: [14715390](https://pubmed.ncbi.nlm.nih.gov/14715390/)]
25. Neubauer AB, Scott SB, Sliwinski MJ, Smyth JM. How was your day? Convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span. *J Pers Soc Psychol*. Jul 2020;119(1):185-203. [FREE Full text] [doi: [10.1037/pspp0000248](https://doi.org/10.1037/pspp0000248)] [Medline: [31070397](https://pubmed.ncbi.nlm.nih.gov/31070397/)]
26. Zurbriggen CL, Jendryczko D, Nussbeck FW. Rosy or blue? Change in recall bias of students' affective experiences during early adolescence. *Emotion*. Dec 2021;21(8):1637-1649. [doi: [10.1037/emo0001031](https://doi.org/10.1037/emo0001031)] [Medline: [34928636](https://pubmed.ncbi.nlm.nih.gov/34928636/)]
27. Chajut E, Caspi A, Chen R, Hod M, Ariely D. In pain thou shalt bring forth children: the peak-and-end rule in recall of labor pain. *Psychol Sci*. Dec 2014;25(12):2266-2271. [doi: [10.1177/0956797614551004](https://doi.org/10.1177/0956797614551004)] [Medline: [25298293](https://pubmed.ncbi.nlm.nih.gov/25298293/)]
28. Kahneman D, Fredrickson BL, Schreier CA, Redelmeier DA. When more pain is preferred to less: adding a better end. *Psychol Sci*. Nov 01, 1993;4(6):401-405. [doi: [10.1111/j.1467-9280.1993.tb00589.x](https://doi.org/10.1111/j.1467-9280.1993.tb00589.x)]
29. Redelmeier DA, Katz J, Kahneman D. Memories of colonoscopy: a randomized trial. *Pain*. Jul 2003;104(1-2):187-194. [doi: [10.1016/s0304-3959\(03\)00003-4](https://doi.org/10.1016/s0304-3959(03)00003-4)] [Medline: [12855328](https://pubmed.ncbi.nlm.nih.gov/12855328/)]
30. Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain*. Jul 1996;66(1):3-8. [doi: [10.1016/0304-3959\(96\)02994-6](https://doi.org/10.1016/0304-3959(96)02994-6)] [Medline: [8857625](https://pubmed.ncbi.nlm.nih.gov/8857625/)]
31. Conner TS, Barrett LF. Trends in ambulatory self-report: the role of momentary experience in psychosomatic medicine. *Psychosom Med*. May 2012;74(4):327-337. [FREE Full text] [doi: [10.1097/PSY.0b013e3182546f18](https://doi.org/10.1097/PSY.0b013e3182546f18)] [Medline: [22582330](https://pubmed.ncbi.nlm.nih.gov/22582330/)]

32. Miron-Shatz T, Stone A, Kahneman D. Memories of yesterday's emotions: does the valence of experience affect the memory-experience gap? *Emotion*. Dec 2009;9(6):885-891. [doi: [10.1037/a0017823](https://doi.org/10.1037/a0017823)] [Medline: [20001131](https://pubmed.ncbi.nlm.nih.gov/20001131/)]
33. Faul L, LaBar KS. Mood-congruent memory revisited. *Psychol Rev*. Dec 2023;130(6):1421-1456. [FREE Full text] [doi: [10.1037/rev0000394](https://doi.org/10.1037/rev0000394)] [Medline: [36201828](https://pubmed.ncbi.nlm.nih.gov/36201828/)]
34. Robinson MD, Clore GL. Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *J Pers Soc Psychol*. 2002;83(1):198-215. [doi: [10.1037//0022-3514.83.1.198](https://doi.org/10.1037//0022-3514.83.1.198)]
35. Tulving E. Episodic memory: from mind to brain. *Annu Rev Psychol*. Feb 2002;53(1):1-25. [doi: [10.1146/annurev.psych.53.100901.135114](https://doi.org/10.1146/annurev.psych.53.100901.135114)] [Medline: [11752477](https://pubmed.ncbi.nlm.nih.gov/11752477/)]
36. Geng X, Chen Z, Lam W, Zheng Q. Hedonic evaluation over short and long retention intervals: the mechanism of the peak-end rule. *J Behav Decis Mak*. May 31, 2013;26(3):225-236. [doi: [10.1002/bdm.1755](https://doi.org/10.1002/bdm.1755)]
37. Walentynowicz M, Schneider S, Stone AA. The effects of time frames on self-report. *PLoS One*. Aug 9, 2018;13(8):e0201655. [FREE Full text] [doi: [10.1371/journal.pone.0201655](https://doi.org/10.1371/journal.pone.0201655)] [Medline: [30092034](https://pubmed.ncbi.nlm.nih.gov/30092034/)]
38. Jaso BA, Kraus NI, Heller AS. Identification of careless responding in ecological momentary assessment research: From posthoc analyses to real-time data monitoring. *Psychol Methods*. Dec 16, 2022;27(6):958-981. [doi: [10.1037/met0000312](https://doi.org/10.1037/met0000312)] [Medline: [34582244](https://pubmed.ncbi.nlm.nih.gov/34582244/)]
39. Shrout PE, Stadler G, Lane SP, McClure MJ, Jackson GL, Clavél FD, et al. Initial elevation bias in subjective reports. *Proc Natl Acad Sci U S A*. Jan 02, 2018;115(1):E15-E23. [FREE Full text] [doi: [10.1073/pnas.1712277115](https://doi.org/10.1073/pnas.1712277115)] [Medline: [29255039](https://pubmed.ncbi.nlm.nih.gov/29255039/)]
40. Cerino ES, Schneider S, Stone AA, Sliwinski MJ, Mogle J, Smyth J. Little evidence for consistent initial elevation bias in self-reported momentary affect: a coordinated analysis of ecological momentary assessment studies. *Psychol Assess*. May 2022;34(5):467-482. [FREE Full text] [doi: [10.1037/pas0001108](https://doi.org/10.1037/pas0001108)] [Medline: [35175074](https://pubmed.ncbi.nlm.nih.gov/35175074/)]
41. Schneider S, Hernandez R, Junghaenel DU, Orriens B, Lee P, Stone AA. Response times in Ecological Momentary Assessment (EMA): shedding light on the response process with a drift diffusion model. *Curr Psychol*. May 27, 2023;43(7):5868-5886. [doi: [10.1007/S12144-023-04773-0](https://doi.org/10.1007/S12144-023-04773-0)]
42. Hernandez R, Schneider S, Pinkham AE, Depp CA, Ackerman R, Pyatak EA, et al. Comparisons of self-report with objective measurements suggest faster responding but little change in response quality over time in ecological momentary assessment studies. *Assessment*. May 18, 2025;32(3):335-355. [doi: [10.1177/10731911241245793](https://doi.org/10.1177/10731911241245793)] [Medline: [38634454](https://pubmed.ncbi.nlm.nih.gov/38634454/)]
43. Borsboom D, Deserno MK, Rhemtulla M, Epskamp S, Fried EI, McNally RJ, et al. Network analysis of multivariate data in psychological science. *Nat Rev Methods Primers*. Aug 19, 2021;1(1):58. [doi: [10.1038/S43586-021-00055-W](https://doi.org/10.1038/S43586-021-00055-W)]
44. Kuppens P, Verduyn P. Emotion dynamics. *Curr Opin Psychol*. Oct 2017;17:22-26. [doi: [10.1016/j.copsyc.2017.06.004](https://doi.org/10.1016/j.copsyc.2017.06.004)] [Medline: [28950968](https://pubmed.ncbi.nlm.nih.gov/28950968/)]
45. Fisher H, Fatimah H, Pidvirny K, Brown H, Balkind E, Pastro B, et al. Affect dynamics in adolescent depression: are all equilibria worth returning to? *Emotion* (Forthcoming). May 26, 2025. [doi: [10.1037/emo0001545](https://doi.org/10.1037/emo0001545)] [Medline: [40424153](https://pubmed.ncbi.nlm.nih.gov/40424153/)]
46. McNeish D, MacKinnon DP. Intensive longitudinal mediation in Mplus. *Psychol Methods*. May 2025;30(2):393-415. [doi: [10.1037/met0000536](https://doi.org/10.1037/met0000536)] [Medline: [36548080](https://pubmed.ncbi.nlm.nih.gov/36548080/)]
47. Dejonckheere E, Mestdagh M. On the signal-to-noise ratio in real-life emotional time series. In: Waugh CE, Kuppens P, editors. *Affect Dynamics*. Cham, Switzerland: Springer; 2021:131-152.
48. Hilt LM, Swords CM, Webb CA. Randomized controlled trial of a mindfulness mobile application for ruminative adolescents. *J Clin Child Adolesc Psychol*. Jan 09, 2025;54(1):99-112. [FREE Full text] [doi: [10.1080/15374416.2022.2158840](https://doi.org/10.1080/15374416.2022.2158840)] [Medline: [36622879](https://pubmed.ncbi.nlm.nih.gov/36622879/)]
49. Hilt LM, Vachhani SS, Vaghiasa NV, Swords CM, Sage AJ. Brief app-based mood monitoring and mindfulness intervention for first-year college students: a randomized controlled trial. *J Psychopathol Behav Assess*. Feb 21, 2025;47(1):23. [doi: [10.1007/S10862-025-10196-X](https://doi.org/10.1007/S10862-025-10196-X)]
50. Choi SW, Schalet B, Cook KF, Cella D. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. *Psychol Assess*. Jul 2014;26(2):513-527. [FREE Full text] [doi: [10.1037/a0035768](https://doi.org/10.1037/a0035768)] [Medline: [24548149](https://pubmed.ncbi.nlm.nih.gov/24548149/)]
51. Schalet BD, Cook KF, Choi SW, Cella D. Establishing a common metric for self-reported anxiety: linking the MASQ, PANAS, and GAD-7 to PROMIS Anxiety. *J Anxiety Disord*. Jan 2014;28(1):88-96. [FREE Full text] [doi: [10.1016/j.janxdis.2013.11.006](https://doi.org/10.1016/j.janxdis.2013.11.006)] [Medline: [24508596](https://pubmed.ncbi.nlm.nih.gov/24508596/)]
52. Healthy minds program app dosage. National Library of Medicine. URL: <https://clinicaltrials.gov/study/NCT05229406> [accessed 2025-05-29]
53. Healthy minds program dosage study: feasibility and acceptability. Open Science Framework. URL: https://osf.io/fszvj?view_only=9cb1b9e67cc042f9bc7a0309e94b2f52 [accessed 2025-05-29]
54. Aalto M, Alho H, Halme JT, Seppä K. AUDIT and its abbreviated versions in detecting heavy and binge drinking in a general population survey. *Drug Alcohol Depend*. Jul 01, 2009;103(1-2):25-29. [doi: [10.1016/j.drugalcdep.2009.02.013](https://doi.org/10.1016/j.drugalcdep.2009.02.013)] [Medline: [19395203](https://pubmed.ncbi.nlm.nih.gov/19395203/)]
55. Dahl CJ, Wilson-Mendenhall CD, Davidson RJ. The plasticity of well-being: a training-based framework for the cultivation of human flourishing. *Proc Natl Acad Sci U S A*. Dec 22, 2020;117(51):32197-32206. [FREE Full text] [doi: [10.1073/pnas.2014859117](https://doi.org/10.1073/pnas.2014859117)] [Medline: [33288719](https://pubmed.ncbi.nlm.nih.gov/33288719/)]

56. Goldberg SB, Imhoff-Smith T, Bolt DM, Wilson-Mendenhall CD, Dahl CJ, Davidson RJ, et al. Testing the efficacy of a multicomponent, self-guided, smartphone-based meditation app: three-armed randomized controlled trial. *JMIR Ment Health*. Dec 27, 2020;7(11):e23825. [FREE Full text] [doi: [10.2196/23825](https://doi.org/10.2196/23825)] [Medline: [33245288](https://pubmed.ncbi.nlm.nih.gov/33245288/)]
57. Hirshberg MJ, Frye C, Dahl CJ, Riordan KM, Vack NJ, Sachs J, et al. A randomized controlled trial of a smartphone-based well-being training in public school system employees during the COVID-19 pandemic. *J Educ Psychol*. Dec 2022;114(8):1895-1911. [FREE Full text] [doi: [10.1037/edu0000739](https://doi.org/10.1037/edu0000739)] [Medline: [36387982](https://pubmed.ncbi.nlm.nih.gov/36387982/)]
58. Kovacs M. The Children's Depression Inventory (CDI) manual. Multi-Health Systems. 1992. URL: <https://www.scienceopen.com/book?vid=761e0bdc-c7a2-44a1-9730-c8aa4bed7424> [accessed 2025-05-29]
59. Craighead WE, Curry JF, Ilardi S. Relationship of children's depression inventory factors to major depression among adolescents. *Psychol Assess*. 1995;7(2):171-176. [doi: [10.1037//1040-3590.7.2.171](https://doi.org/10.1037//1040-3590.7.2.171)]
60. Klein DN, Dougherty LR, Olino TM. Toward guidelines for evidence-based assessment of depression in children and adolescents. *J Clin Child Adolesc Psychol*. Oct 2005;34(3):412-432. [doi: [10.1207/s15374424jccp3403_3](https://doi.org/10.1207/s15374424jccp3403_3)] [Medline: [16026212](https://pubmed.ncbi.nlm.nih.gov/16026212/)]
61. Storch EA, Roberti JW, Roth DA. Factor structure, concurrent validity, and internal consistency of the beck depression inventory-second edition in a sample of college students. *Depress Anxiety*. 2004;19(3):187-189. [doi: [10.1002/da.20002](https://doi.org/10.1002/da.20002)] [Medline: [15129421](https://pubmed.ncbi.nlm.nih.gov/15129421/)]
62. Pilkonis P, Choi S, Reise S, Stover A, Riley W, Cella D, et al. PROMIS Cooperative Group. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. *Assessment*. Oct 2011;18(3):263-283. [FREE Full text] [doi: [10.1177/1073191111411667](https://doi.org/10.1177/1073191111411667)] [Medline: [21697139](https://pubmed.ncbi.nlm.nih.gov/21697139/)]
63. Nolte S, Coon C, Hudgens S, Verdam MG. Psychometric evaluation of the PROMIS® depression item bank: an illustration of classical test theory methods. *J Patient Rep Outcomes*. Jul 30, 2019;3(1):46. [FREE Full text] [doi: [10.1186/s41687-019-0127-0](https://doi.org/10.1186/s41687-019-0127-0)] [Medline: [31359212](https://pubmed.ncbi.nlm.nih.gov/31359212/)]
64. Nolen-Hoeksema S, Wisco BE, Lyubomirsky S. Rethinking rumination. *Perspect Psychol Sci*. Oct 2008;3(5):400-424. [FREE Full text] [doi: [10.1111/j.1745-6924.2008.00088.x](https://doi.org/10.1111/j.1745-6924.2008.00088.x)] [Medline: [26158958](https://pubmed.ncbi.nlm.nih.gov/26158958/)]
65. Nolen-Hoeksema S, Morrow J. A prospective study of depression and posttraumatic stress symptoms after a natural disaster: the 1989 Loma Prieta Earthquake. *J Pers Soc Psychol*. Jul 1991;61(1):115-121. [doi: [10.1037//0022-3514.61.1.115](https://doi.org/10.1037//0022-3514.61.1.115)] [Medline: [1890582](https://pubmed.ncbi.nlm.nih.gov/1890582/)]
66. Roelofs J, Muris P, Huibers M, Peeters F, Arntz A. On the measurement of rumination: a psychometric evaluation of the ruminative response scale and the rumination on sadness scale in undergraduates. *J Behav Ther Exp Psychiatry*. Dec 2006;37(4):299-313. [doi: [10.1016/j.jbtep.2006.03.002](https://doi.org/10.1016/j.jbtep.2006.03.002)] [Medline: [16678121](https://pubmed.ncbi.nlm.nih.gov/16678121/)]
67. Ehring T, Zetsche U, Weidacker K, Wahl K, Schönfeld S, Ehlers A. The Perseverative Thinking Questionnaire (PTQ): validation of a content-independent measure of repetitive negative thinking. *J Behav Ther Exp Psychiatry*. Jul 2011;42(2):225-232. [FREE Full text] [doi: [10.1016/j.jbtep.2010.12.003](https://doi.org/10.1016/j.jbtep.2010.12.003)] [Medline: [21315886](https://pubmed.ncbi.nlm.nih.gov/21315886/)]
68. Hilt LM, Pollak SD. Getting out of rumination: comparison of three brief interventions in a sample of youth. *J Abnorm Child Psychol*. Oct 22, 2012;40(7):1157-1165. [FREE Full text] [doi: [10.1007/s10802-012-9638-3](https://doi.org/10.1007/s10802-012-9638-3)] [Medline: [22527609](https://pubmed.ncbi.nlm.nih.gov/22527609/)]
69. Moberly NJ, Watkins ER. Ruminative self-focus and negative affect: an experience sampling study. *J Abnorm Psychol*. May 2008;117(2):314-323. [FREE Full text] [doi: [10.1037/0021-843X.117.2.314](https://doi.org/10.1037/0021-843X.117.2.314)] [Medline: [18489207](https://pubmed.ncbi.nlm.nih.gov/18489207/)]
70. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University. URL: <https://cran.r-project.org/web/packages/psych/index.html> [accessed 2025-08-06]
71. Ruscio AM, Gentes EL, Jones JD, Hallion LS, Coleman ES, Swendsen J. Rumination predicts heightened responding to stressful life events in major depressive disorder and generalized anxiety disorder. *J Abnorm Psychol*. Mar 2015;124(1):17-26. [FREE Full text] [doi: [10.1037/abn0000025](https://doi.org/10.1037/abn0000025)] [Medline: [25688429](https://pubmed.ncbi.nlm.nih.gov/25688429/)]
72. Webb CA, Israel ES, Belleau E, Appleman L, Forbes EE, Pizzagalli DA. Mind-wandering in adolescents predicts worse affect and is linked to aberrant default mode network-salience network connectivity. *J Am Acad Child Adolesc Psychiatry*. Mar 2021;60(3):377-387. [FREE Full text] [doi: [10.1016/j.jaac.2020.03.010](https://doi.org/10.1016/j.jaac.2020.03.010)] [Medline: [32553785](https://pubmed.ncbi.nlm.nih.gov/32553785/)]
73. Williams RH, Zimmerman DW. The comparative reliability of simple and residualized difference scores. *J Exp Educ*. Jan 28, 2015;51(2):94-97. [doi: [10.1080/00220973.1982.11011846](https://doi.org/10.1080/00220973.1982.11011846)]
74. Homepage. National Center on Intensive Intervention. 2014. URL: https://intensiveintervention.org/sites/default/files/APM_FAQs_2014.pdf [accessed 2024-05-06]
75. Van Norman ER, Parker DC. A comparison of split-half and multilevel methods to assess the reliability of progress monitoring outcomes. *J Psychoeduc Assess*. Mar 08, 2017;36(6):616-627. [doi: [10.1177/0734282917696936](https://doi.org/10.1177/0734282917696936)]
76. VanDerHeyden AM, Burns MK. Examination of the utility of various measures of mathematics proficiency. *Assess Eff Interv*. Sep 01, 2008;33(4):215-224. [doi: [10.1177/1534508407313482](https://doi.org/10.1177/1534508407313482)]
77. Raudenbush SW, Bryk AS. Hierarchical Linear Models: Applications and Data Analysis Methods. Thousand Oaks, CA. Sage Publication; 2002.
78. Liu S, Kuppens P, Bringmann L. On the use of empirical bayes estimates as measures of individual traits. *Assessment*. May 31, 2021;28(3):845-857. [doi: [10.1177/1073191119885019](https://doi.org/10.1177/1073191119885019)] [Medline: [31672023](https://pubmed.ncbi.nlm.nih.gov/31672023/)]
79. Harrell F. Statistical errors in the medical literature. *Statistical Thinking*. URL: <https://www.fharrell.com/post/errmed/#change> [accessed 2025-06-19]

80. Feingold A. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychol Methods*. Mar 2009;14(1):43-53. [FREE Full text] [doi: [10.1037/a0014699](https://doi.org/10.1037/a0014699)] [Medline: [19271847](https://pubmed.ncbi.nlm.nih.gov/19271847/)]
81. Hilt LM, Sladek MR, Doane LD, Stroud CB. Daily and trait rumination: diurnal cortisol patterns in adolescent girls. *Cogn Emot*. Dec 2017;31(8):1757-1767. [FREE Full text] [doi: [10.1080/02699931.2016.1262332](https://doi.org/10.1080/02699931.2016.1262332)] [Medline: [27922343](https://pubmed.ncbi.nlm.nih.gov/27922343/)]
82. Crocker LM, Algina J. Introduction to Classical and Modern Test Theory. London, UK. Holt, Rinehart, & Winston; 1986.
83. Cohen J. Statistical Power Analysis for the Behavioral Science. 2 nd edition. Hillsdale, NJ. Routledge; 1988.
84. Brose A, Schmiedek F, Gerstorf D, Voelkle MC. The measurement of within-person affect variation. *Emotion*. Jul 2020;20(4):677-699. [doi: [10.1037/emo0000583](https://doi.org/10.1037/emo0000583)] [Medline: [31008618](https://pubmed.ncbi.nlm.nih.gov/31008618/)]
85. Cronbach LJ, Furby L. How we should measure "change": or should we? *Psychol Bull*. Jul 1970;74(1):68-80. [doi: [10.1037/h0029382](https://doi.org/10.1037/h0029382)]
86. Willett JB. Some results on reliability for the longitudinal measurement of change: implications for the design of studies of individual growth. *Educ Psychol Meas*. Sep 01, 1989;49(3):587-602. [doi: [10.1177/001316448904900309](https://doi.org/10.1177/001316448904900309)]
87. Williams RH, Zimmerman DW. Are simple gain scores obsolete? *Appl Psychol Meas*. 1996;20(1):59-69. [doi: [10.1177/0146621696020001](https://doi.org/10.1177/0146621696020001)]
88. Cranford JA, Shrout PE, Iida M, Rafaeli E, Yip T, Bolger N. A procedure for evaluating sensitivity to within-person change: can mood measures in diary studies detect change reliably? *Pers Soc Psychol Bull*. Jul 2006;32(7):917-929. [FREE Full text] [doi: [10.1177/0146167206287721](https://doi.org/10.1177/0146167206287721)] [Medline: [16738025](https://pubmed.ncbi.nlm.nih.gov/16738025/)]
89. Dejonckheere E, Demeyer F, Geusens B, Piot M, Tuerlinckx F, Verdonck S, et al. Assessing the reliability of single-item momentary affective measurements in experience sampling. *Psychol Assess*. Dec 2022;34(12):1138-1154. [doi: [10.1037/pas0001178](https://doi.org/10.1037/pas0001178)] [Medline: [36074609](https://pubmed.ncbi.nlm.nih.gov/36074609/)]
90. Haney AM, Fleming MN, Wycoff AM, Griffin SA, Trull TJ. Measuring affect in daily life: a multilevel psychometric evaluation of the PANAS-X across four ecological momentary assessment samples. *Psychol Assess*. Jul 2023;35(6):469-483. [FREE Full text] [doi: [10.1037/pas0001231](https://doi.org/10.1037/pas0001231)] [Medline: [36931821](https://pubmed.ncbi.nlm.nih.gov/36931821/)]
91. Shrout PE, Lane SP. Psychometrics. In: Mehl MR, Conner TS, editors. *Handbook of Research Methods for Studying Daily Life*. New York, NY. The Guilford Press; 2012:302-320.
92. Parsons S, Kruijt AW, Fox E. Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Adv Methods Pract Psychol Sci*. Nov 06, 2019;2(4):378-395. [doi: [10.1177/2515245919879695](https://doi.org/10.1177/2515245919879695)]
93. Bolger N, Laurenceau JP. *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research*. New York, NY. Guilford Press; 2013.
94. van Berkel N, Goncalves J, Hosio S, Sarsenbayeva Z, Velloso E, Kostakos V. Overcoming compliance bias in self-report studies: a cross-study analysis. *Int J Hum Comput Stud*. Feb 2020;134:1-12. [doi: [10.1016/j.ijhcs.2019.10.003](https://doi.org/10.1016/j.ijhcs.2019.10.003)]
95. Loas G, Monestes JL, Ingelaere A, Noisette C, Herbener E. Stability and relationships between trait or state anhedonia and schizophrenic symptoms in schizophrenia: a 13-year follow-up study. *Psychiatry Res*. May 30, 2009;166(2-3):132-140. [doi: [10.1016/j.psychres.2008.02.010](https://doi.org/10.1016/j.psychres.2008.02.010)] [Medline: [19272653](https://pubmed.ncbi.nlm.nih.gov/19272653/)]
96. Anestis MD, Selby EA, Crosby RD, Wonderlich SA, Engel SG, Joiner TE. A comparison of retrospective self-report versus ecological momentary assessment measures of affective lability in the examination of its relationship with bulimic symptomatology. *Behav Res Ther*. Jul 2010;48(7):607-613. [FREE Full text] [doi: [10.1016/j.brat.2010.03.012](https://doi.org/10.1016/j.brat.2010.03.012)] [Medline: [20392437](https://pubmed.ncbi.nlm.nih.gov/20392437/)]
97. Solhan MB, Trull TJ, Jahng S, Wood PK. Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. *Psychol Assess*. Oct 2009;21(3):425-436. [FREE Full text] [doi: [10.1037/a0016869](https://doi.org/10.1037/a0016869)] [Medline: [19719353](https://pubmed.ncbi.nlm.nih.gov/19719353/)]
98. Rauthmann JF, Horstmann KT, Sherman RA. Do self-reported traits and aggregated states capture the same thing? A nomological perspective on trait-state homomorphy. *Soc Psychol Personal Sci*. Jun 19, 2018;10(5):596-611. [doi: [10.1177/1948550618774772](https://doi.org/10.1177/1948550618774772)]
99. Ringwald WR, Manuck SB, Marsland AL, Wright AG. Psychometric evaluation of a big five personality state scale for intensive longitudinal studies. *Assessment*. Oct 2022;29(6):1301-1319. [FREE Full text] [doi: [10.1177/10731911211008254](https://doi.org/10.1177/10731911211008254)] [Medline: [33949209](https://pubmed.ncbi.nlm.nih.gov/33949209/)]
100. Kahneman D. *Thinking, Fast and Slow*. New York, NY. Farrar, Straus & Giroux; 2011.
101. Augustine AA, Larsen RJ. Is a trait really the mean of states? *J Individ Differ*. Jan 2012;33(3):131-137. [doi: [10.1027/1614-0001/a000083](https://doi.org/10.1027/1614-0001/a000083)]
102. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol*. Apr 01, 2008;4(1):1-32. [doi: [10.1146/annurev.clinpsy.3.022806.091415](https://doi.org/10.1146/annurev.clinpsy.3.022806.091415)] [Medline: [18509902](https://pubmed.ncbi.nlm.nih.gov/18509902/)]
103. Piccirillo ML, Rodebaugh TL. Foundations of idiographic methods in psychology and applications for psychotherapy. *Clin Psychol Rev*. Jul 2019;71:90-100. [FREE Full text] [doi: [10.1016/j.cpr.2019.01.002](https://doi.org/10.1016/j.cpr.2019.01.002)] [Medline: [30665765](https://pubmed.ncbi.nlm.nih.gov/30665765/)]
104. Kazdin AE. Mediators and mechanisms of change in psychotherapy research. *Annu Rev Clin Psychol*. Apr 01, 2007;3(1):1-27. [doi: [10.1146/annurev.clinpsy.3.022806.091432](https://doi.org/10.1146/annurev.clinpsy.3.022806.091432)] [Medline: [17716046](https://pubmed.ncbi.nlm.nih.gov/17716046/)]
105. Shahar E. Evaluating the effect of change on change: a different viewpoint. *J Eval Clin Pract*. Mar 12, 2009;15(1):204-207. [doi: [10.1111/j.1365-2753.2008.00983.x](https://doi.org/10.1111/j.1365-2753.2008.00983.x)] [Medline: [19239603](https://pubmed.ncbi.nlm.nih.gov/19239603/)]

106. Fried EI, van Borkulo CD, Epskamp S, Schoevers R, Tuerlinckx F, Borsboom D. Measuring depression over time . . . or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychol Assess*. Dec 2016;28(11):1354-1367. [FREE Full text] [doi: [10.1037/pas0000275](https://doi.org/10.1037/pas0000275)] [Medline: [26821198](https://pubmed.ncbi.nlm.nih.gov/26821198/)]
107. McNeish D, Mackinnon DP, Marsch LA, Poldrack RA. Measurement in intensive longitudinal data. *Struct Equ Modeling*. May 24, 2021;28(5):807-822. [FREE Full text] [doi: [10.1080/10705511.2021.1915788](https://doi.org/10.1080/10705511.2021.1915788)] [Medline: [34737528](https://pubmed.ncbi.nlm.nih.gov/34737528/)]
108. Van Dam NT, Earleywine M, Danoff-Burg S. Differential item function across meditators and non-meditators on the five facet mindfulness questionnaire. *Pers Individ Dif*. Oct 2009;47(5):516-521. [doi: [10.1016/j.paid.2009.05.005](https://doi.org/10.1016/j.paid.2009.05.005)]
109. Chang EC, Tsai W, Sanna LJ. Examining the relations between rumination and adjustment: do ethnic differences exist between Asian and European Americans? *Asian Am J Psychol*. Mar 2010;1(1):46-56. [doi: [10.1037/A0018821](https://doi.org/10.1037/A0018821)]
110. Bandalos DL. *Measurement Theory and Applications for the Social Sciences*. New York, NY. The Guilford Press; 2018.
111. Collins LM. Is reliability obsolete? A commentary on "are simple gain scores obsolete?". *Appl Psychol Meas*. Sep 01, 1996;20(3):289-292. [doi: [10.1177/014662169602000308](https://doi.org/10.1177/014662169602000308)]
112. Thomas DR, Zumbo BD. Difference scores from the point of view of reliability and repeated-measures ANOVA. *Educ Psychol Meas*. Jun 02, 2011;72(1):37-43. [doi: [10.1177/0013164411409929](https://doi.org/10.1177/0013164411409929)]
113. Trafimow D. A defense against the alleged unreliability of difference scores. *Cogent Math*. Jul 17, 2015;2(1):1064626. [doi: [10.1080/23311835.2015.1064626](https://doi.org/10.1080/23311835.2015.1064626)]
114. Are ecological momentary assessment measures of intervention change worth the trouble? Evaluation in four digital mental health trials. *Open Science Framework*. URL: <https://osf.io/2bnmk/> [accessed 2024-05-29]

Abbreviations

BDI-II: Beck Depression Inventory–II
CARE: Child and Adolescent Research in Emotion
CDI: Children's Depression Inventory
CRSQ: Children's Response Styles Questionnaire
EMA: ecological momentary assessment
HMP: Healthy Minds Program
LMM: linear mixed-effects model
MLM: multilevel model
OLS: ordinary least squares
PROMIS: Patient-Reported Outcomes Measurement Information System
PTQ: Perseverative Thinking Questionnaire
RRS: Ruminative Response Scale
RSQ: Response Styles Questionnaire
SB: Spearman-Brown

Edited by T de Azevedo Cardoso; submitted 26.11.24; peer-reviewed by CAL Bean, NH Zainal, MR Frumkin; comments to author 13.06.25; revised version received 23.06.25; accepted 16.07.25; published 09.09.25

Please cite as:

Webb CA, Hilt LM, Swords CM, Bolt DM, Fisher H, Goldberg SB
Ecological Momentary Assessment as a Measure of Intervention Change: Evaluation in 4 Digital Mental Health Trials
J Med Internet Res 2025;27:e69297
URL: <https://www.jmir.org/2025/1/e69297>
doi: [10.2196/69297](https://doi.org/10.2196/69297)
PMID:

©Christian A Webb, Lori M Hilt, Caroline M Swords, Daniel M Bolt, Hadar Fisher, Simon B Goldberg. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 09.09.2025. 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 (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.