

Original Paper

# An Early-Stage Digital Therapeutic Intervention to Enhance Affective Response During Physical Activity Among Adults With Overweight or Obesity: Benchmark-Driven Formative Testing Study

Rachel Crosley-Lyons<sup>1</sup>, MS; Lori Hatzinger<sup>1</sup>, MS; Micaela Hewus<sup>1</sup>, MPH; Wei-Lin Wang<sup>1</sup>, PhD; Delfien Van Dyck<sup>2</sup>, PhD; Jimi Huh<sup>1</sup>, PhD; Eric Hekler<sup>3,4</sup>, PhD; Genevieve F Dunton<sup>1,5</sup>, MPH, PhD

<sup>1</sup>Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

<sup>2</sup>Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium

<sup>3</sup>Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, San Diego, CA, United States

<sup>4</sup>Center for Wireless & Population Health Systems, University of California, San Diego, San Diego, CA, United States

<sup>5</sup>Department of Psychology, University of Southern California, Los Angeles, CA, United States

**Corresponding Author:**

Rachel Crosley-Lyons, MS

Department of Population and Public Health Sciences

Keck School of Medicine

University of Southern California

1845 N Soto Street

Los Angeles, CA, 90032

United States

Phone: 1 3175187670

Email: [rclyons@usc.edu](mailto:rclyons@usc.edu)

## Abstract

**Background:** Mobile device-enabled interventions known as digital therapeutics (DTx) are increasingly used to prevent chronic disease by targeting psychological and behavioral processes. Individuals' unique experiences while receiving DTx comprise real-world evidence (RWE) for evaluating DTx performance. An emerging strategy for early-stage DTx formative work uses small sample sizes, which facilitate efficient iteration and agile learning, while evaluating performance against descriptive benchmarks defined a priori, therefore minimizing the risk for confirmation bias. This study test benchmarks from the DTx RWE framework to formatively evaluate a novel DTx (the eMOTION study) to enhance affective response (ie, how people feel) during physical activity (PA).

**Objective:** This study aimed to determine whether the eMOTION DTx met a priori benchmarks for safety (<1% of participants report an adverse event), plausibility (≥51% of participants experience increased enjoyment in PA), usability (eg, ≥51% of participants report adequate usability), sustainability, feasibility (eg, <70% of participants report dissatisfaction), and equity (equity and accessibility are approximately equal across subgroups).

**Methods:** Participants (N=36; mean age 46, SD 14 years; 20/37, 54% female) underwent stratified random assignment to test one of four DTx versions for 14 days (n=9 each): (1) intensity PA goals, (2) affect PA goals with type and context recommendations, (3) affect PA goals with savoring exercises, and (4) affect PA goals with type, context, and savoring. Participants completed daily intervention sessions, asking them to focus on achieving a target heart rate (intensity) or feeling good (affect) during PA. Smartwatches were used to track PA and answer ecological momentary assessment (EMA) questions about how they felt during PA. Performance toward benchmarks was primarily assessed via official Institutional Review Board reporting channels (safety), interviews (plausibility, accessibility, and usability), and questionnaires (System Usability Scale [usability], Delighted-Terrible Scale [sustainability and feasibility], and equity).

**Results:** The eMOTION DTx versions exceeded all a priori safety, plausibility, accessibility, usability, sustainability, feasibility, and equity thresholds. For safety, no adverse events were reported. Regarding plausibility, more than half of the participants who received affect goals reported increased PA enjoyment at the end of the study. Moreover, 64%-72% (23-26 out of 36) of participants rated the DTx at or above the standard System Usability Scale cutoff point for acceptable usability. More than 60% (22/36) of

participants reported satisfaction with all DTx components, supporting DTx sustainability and feasibility. Finally, there was evidence for equity, with plausibility and accessibility comparable across sex, race, ethnicity, income, age, BMI, mobility, and physical constraint subgroups.

**Conclusions:** Since DTx RWE Framework benchmarks for safety, plausibility, accessibility, usability, sustainability, feasibility, and equity were largely met, the eMOTION Study DTx is ready for a full-scale efficacy trial to refine the DTx and optimize efficiency and feasibility. Our approach incorporated transparent decision-making to generate results that are more readily translatable, easily replicable, and reflect current best practices in the field of DTx.

**Trial Registration:** ClinicalTrials.gov NCT06125964; <https://clinicaltrials.gov/study/NCT06125964>

(*JMIR Hum Factors* 2026;13:e71472) doi: [10.2196/71472](https://doi.org/10.2196/71472)

## KEYWORDS

digital therapeutics; real world evidence; formative study; MOST framework; SMART goals; mHealth; physical activity; affective response

## Introduction

### Digital Therapeutics

Widespread smartphone ownership and expanded wearable device capabilities have steadily facilitated a paradigm shift in health care delivery. The medical system has traditionally used in-person examinations and pharmaceuticals to detect, treat, and manage chronic diseases. However, such diagnostic care is incredibly costly—with global spending projected to reach US \$47 trillion by 2030—and inefficient, given that many noncommunicable diseases are potentially avoidable via behavior modification [1]. While clinicians counsel patients about the importance of a healthy lifestyle for mitigating chronic disease risk, compliance with preventive care recommendations remains low. Digital therapeutics (DTx)—mobile device software systems that deliver evidence-based behavioral intervention packages to treat or prevent disease in real-world environments—are poised to complement and improve upon these efforts. For example, DTx can extend preventive care beyond the clinic by facilitating behavior change in daily life and empowering patients to monitor their own progress. High-resolution digital treatment evaluation data collected by DTx serve as real-world evidence (RWE) that can be used for efficient performance assessments during research and development. Even as forecasts show a DTx market compound annual growth rate of 27.2% until 2030 [2], the transition from DTx research and development to regulatory testing and public release remains disjointed and inefficient.

### Benchmark-Driven Formative Testing as a Rigorous Approach to Early-Stage DTx Development

DTx involves complex sociotechnical systems with multiple interdependent components—user interfaces, behavioral protocols, technical infrastructure, and human-technology interactions—each of which must function adequately for the intervention to succeed. Failure to establish a robust foundation across all these components before initiating large-scale trials substantially increases the risk of type III error, wherein a study is conducted rigorously but fails to test what it intended to test because the intervention itself was inadequately developed or the control condition was poorly specified.

Traditional approaches to early-stage DTx research face a methodological paradox: formative studies must remain small

to enable rapid, resource-efficient learning and iteration; yet, small sample sizes preclude the use of inferential statistics with adequate statistical power. This creates a problematic situation where researchers either (1) inappropriately apply underpowered inferential tests that are likely to produce unreliable results, or (2) rely solely on qualitative interpretation of descriptive results, which introduces substantial risk of confirmation bias, the well-documented human tendency to interpret ambiguous data in ways that confirm preexisting beliefs and expectations.

To address this methodological challenge, leading experts in behavioral intervention science and DTx development have advocated for benchmark-driven formative testing as a rigorous alternative approach. This methodology, articulated in frameworks such as the Obesity-Related Behavioral Intervention Trials (ORBIT) model for behavioral intervention development and Phase II of the DTx RWE Framework, establishes a priori performance thresholds that can be evaluated using purely descriptive statistics while maintaining scientific rigor [3].

### The Logic of Benchmark-Driven Formative Testing

Benchmark-driven formative testing addresses 4 fundamental tensions in early-stage intervention research.

First, the need for efficient learning requires small sample sizes that enable rapid iteration cycles and resource-efficient exploration of intervention components, implementation strategies, and potential adaptations for diverse populations. Large trials are essential for establishing generalizability, but they are inappropriate tools for the formative work of intervention refinement.

Second, the inappropriateness of underpowered inferential statistics in small samples is well established. When sample sizes are insufficient for adequate statistical power,  $P$  values become unreliable indicators of true effects, CIs become too wide to be informative, and the risk of both false positives and false negatives increases substantially. Running inferential statistics on samples of  $N=30-50$  participants typically yields a power of 0.20-0.40 for detecting medium effect sizes—well below the conventional 0.80 standard. In such contexts, null hypothesis significance testing provides an illusion of rigor while actually undermining the validity of conclusions [4,5].

Third, the risk of confirmation bias affects all human judgment, including scientific interpretation. Without prespecified criteria

for success, researchers may unconsciously interpret ambiguous results favorably, seeing “promising trends” or “encouraging patterns” that align with their hypotheses. This risk is particularly acute in small samples where data can appear to support multiple narratives depending on how one frames the analysis.

Fourth, the distinction between statistical and clinical significance is critical but often conflated. A statistically significant finding may have minimal real-world impact, while a clinically meaningful effect may not reach statistical significance in an underpowered study. Effect sizes, minimal clinically important differences (MCIDs), and practical significance must be considered alongside *P* values; yet, traditional null hypothesis testing frameworks prioritize statistical significance over substantive meaningfulness [6].

### How Benchmarks Address These Tensions

A priori establishment of benchmarks addresses all 4 tensions simultaneously.

#### *Enabling Small, Efficient Samples*

Because benchmarks can be evaluated descriptively (eg, “did 51% of participants report increased PA enjoyment?”). They do not require the large samples necessary for adequate statistical power. This enables formative studies to remain small enough for rapid iteration while still providing clear, interpretable evidence about intervention performance.

#### *Avoiding Inappropriate Statistical Tests*

Benchmark evaluation relies on descriptive statistics—percentages, frequencies, and comparisons to predetermined thresholds—rather than inferential tests. This eliminates the problems associated with underpowered hypothesis testing while still providing rigorous evaluation criteria relevant for formative work within small sample sizes.

#### *Minimizing Confirmation Bias*

By establishing success criteria before data collection, benchmark approaches create an objective standard against which results must be evaluated. If the benchmark specifies that  $\geq 51\%$  of participants must report increased PA enjoyment, then 45% represents clear failure to meet the criterion, regardless of how “close” it might feel or how researchers might be tempted to interpret it favorably. The threshold is predetermined and nonnegotiable.

#### *Centering Clinical Meaningfulness*

Benchmarks are explicitly designed to reflect what would constitute meaningful real-world impact, not merely statistical detectability. For example, in our study here, a 51% threshold for plausibility reflects the expectation that an intervention should benefit the majority of recipients—a clinically meaningful standard that is independent of sample size or statistical power considerations. Benchmarks can incorporate established MCIDs, align with clinically significant thresholds from previous literature, and balance what is both meaningful and plausibly achievable given the current state of the science.

### *Benchmark Selection and Justification*

Rigorous benchmark-driven research requires that thresholds be established through a systematic process that balances clinical meaningfulness with plausible achievability based on previous evidence. Given the novelty of this line of work, our team developed benchmarks that have an initial plausibility of balancing both ambition and likelihood of being observed. While these benchmarks were not explicitly based on previous literature—due to the lack of extant work—they represent a potential baseline that others can improve upon in future proof-of-concept trials.

The process of benchmark development should be transparent and documented, ideally through preregistration, to ensure that thresholds are truly a priori rather than post hoc. In our study, benchmarks were established before data collection and documented in our preregistered analysis plan on Open Science Framework [7].

### *Interpreting Benchmark Results*

Critically, results from benchmark-driven formative studies must not be interpreted as generalizable to broader populations. Meeting benchmarks indicates that an intervention has achieved a sufficient foundation to warrant progression to the next development phase, but it does not establish external validity or population-level efficacy or effectiveness. Those determinations require adequately powered randomized controlled trials with representative samples.

Instead, benchmark-driven formative testing serves a specific purpose in the intervention development pipeline: to establish whether all essential components of an intervention function adequately, whether the intervention can be delivered as intended, whether it demonstrates sufficient signal of benefit to justify further investment, and whether potential implementation barriers can be identified and addressed before committing to large-scale testing. This is analogous to Phase I/II trials in pharmaceutical development, which establish safety, dosing, and preliminary efficacy signals before proceeding to Phase III confirmatory trials.

Meeting formative benchmarks provides evidence that an intervention is “ready” for larger-scale testing—that the foundation is sufficiently solid that a properly powered trial can be conducted without high risk of Type III error. Failing to meet benchmarks indicates that further refinement is needed before proceeding, thereby protecting against the waste of resources on large trials of interventions that are not yet adequately developed.

### *Rigor in Formative Research*

The methodological rigor of formative research is achieved not through the application of inferential statistics to underpowered samples, but through:

1. Systematic benchmark establishment grounded in previous literature and clinical meaningfulness.
2. Prespecification and transparent documentation of all benchmarks before data collection.
3. Comprehensive assessment across multiple domains (safety, plausibility, usability, feasibility, and equity).

4. Objective evaluation of whether benchmarks were met using descriptive statistics.
5. Honest interpretation that acknowledges the formative, nongeneralizable nature of findings.
6. Iterative refinement based on benchmark performance before proceeding to larger trials.

This approach represents a methodologically sound framework for early-stage intervention development that avoids the false precision of underpowered inferential testing while maintaining scientific rigor through prespecified, clinically meaningful evaluation criteria. As Czajkowski and colleagues [3] note in their comprehensive guide to behavioral intervention development, early-stage work should prioritize intervention optimization, rather than hypothesis testing—a fundamentally different scientific objective that requires different methodological tools.

By embracing benchmark-driven formative testing, the field can conduct more rigorous early-stage research that provides a robust foundation for subsequent confirmatory trials. Ultimately, to improve the quality and efficiency of the entire intervention research-to-development pipeline, such early-stage work should systematically and iteratively assess DTx using real-world RWE. The DTx RWE Framework recently proposed by Kim and colleagues [8] incorporates RWE-guided decision-making against benchmarks congruent with existing regulatory standards throughout the DTx lifecycle. The Framework's step-by-step process aims to streamline the research and development-regulatory testing-public release process, facilitate greater transparency, and produce DTx that reflect current best practices and are more readily translatable to real-life settings. This has important public health implications, as emerging DTx are poised to collectively reduce health care burden and expand access to high-quality care via tailoring algorithms that deliver personalized and precise multicomponent treatment [9]. This study applied principles from the DTx RWE Framework in conducting benchmark-driven formative testing of an early-stage DTx intervention to enhance affective response during physical activity (PA) among physically inactive adults with overweight or obesity.

## This Study

### Overview

Inadequate PA engagement (ie, <150 min/week of moderate intensity physical activity [MPA]) confers substantial risk for a variety of chronic diseases, including cancer [10]. Despite widespread public health initiatives—and many individuals reporting a desire to become more physically active—less than 5% of adults actually meet the ≥150 minute/week MPA threshold [11]. Interventions targeting social-cognitive factors (eg, self-efficacy, intentions, and beliefs) have had limited success, accounting for only 31% of the variability in PA engagement [12–15]. This PA intention-behavior gap suggests that noncognitive factors might be hijacking people's best intentions to engage in PA [16]. According to the theory of hedonic motivation, humans are naturally drawn to activities that are pleasurable and avoid unpleasant activities [17]. Previous affective responses to a given behavior inform the development of affectively charged motivations to pursue or

avoid the behavior in the future [18]. People with overweight or obesity are more likely than individuals of a healthy weight to experience pronounced negative affect (eg, discomfort) while engaging in PA, therefore decreasing the likelihood they will consistently engage in PA [19].

Factors that are potentially readily modifiable, such as PA intensity, type, and context, have been shown to be significant predictors of the affective response during PA. For example, most people experience pleasure when engaging in low-intensity PA and displeasure while engaging in vigorous-intensity PA [20], but affective responses to MPA are heterogeneous [21]. At moderate intensities, affective responses might be determined by the type (eg, hiking and dancing) or context (eg, being with others and outdoors) of PA [22–25]. Ultimately, affective responses vary inter- and intrapersonally and impact future PA behavior—but whether they can be experimentally manipulated to promote PA engagement in the real world remains unknown [26,27].

### Developing a DTx Solution

Therefore, an existing DTx called MyDayPlan was modified to experimentally manipulate affective mechanisms associated with MPA among physically inactive adults with overweight or obesity. MyDayPlan has previously increased PA engagement in similar populations by leveraging self-regulation strategies posed by the Health Action Process Approach [28]. Specifically, smartphone sessions provide daily PA goals and prompt participants to plan how, where, and when they will accomplish the goal (action planning), identify potential barriers and solutions (coping planning), and monitor their goal progress (self-monitoring) [29]. In its original form, MyDayPlan provided goals focused on sustaining a certain intensity of PA (ie, “intensity-based goals”). The eMOTION Study incorporated affect-based PA goals and enhancements to modify affective mechanisms related to PA into MyDayPlan's existing self-regulatory framework. New affect-based goals asked participants to engage in PA types and contexts that would allow them to experience positive affect (eg, enjoyment) during PA. We hypothesized that participants who received affect-based goals would report more positive affective experiences during PA than participants who received intensity-based goals, which asked participants to attain a target heart rate. Iterative internal testing among research team members was performed to achieve a minimum viable product with a functioning interface, error-free programming, and a reasonable time burden.

### Benchmark-Driven Formative Testing

Next, following guidance from Phase II of the DTx RWE framework, SMART (specific, measurable, actionable, realistic, and timely [or time-bound]) benchmarks were defined (Textbox 1). Benchmarks required the DTx to reach safety, plausibility, accessibility, usability, sustainability, feasibility, and equity standards during a formative study before initiating a resource-intensive Comparison Study to test efficacy. Broadly, benchmarks ensured that the DTx was unlikely to cause harm (ie, safety); the affect-based goals were associated with greater PA enjoyment compared to the intensity-based control (ie, plausibility); and the DTx was easy to use (ie, accessibility and usability); and practical for study staff and participants to sustain

over time (ie, feasibility and sustainability). We also explored whether DTx plausibility and accessibility were approximately equal for all participants, regardless of individual-level traits (ie, equity; this benchmark was exploratory and not statistically

powered). This paper summarizes results from benchmark-driven formative testing of the eMOTION DTx by measuring performance against these a priori established benchmarks.

**Textbox 1.** eMOTION Formative Study SMART (specific, measurable, actionable, realistic, and timely [or time-bound]) benchmarks.

#### Safety:

- Only <1% of Formative Study participants will experience an adverse event.

#### Plausibility:

- The majority of participants (ie,  $\geq 51\%$ ) receiving affect-based goals will report a perceived increase in enjoyment of physical activity at the end of the study.

#### Accessibility and usability:

- The daily intervention sessions and Fitbit smartwatch features will receive an average score at or above the System Usability Scale cutoff point ( $\geq 68$ ).
- The majority of participants (ie,  $\geq 51\%$ ) will report that they are able to (1) read the wording of the survey questions, (2) understand the wording of the survey questions, (3) select their answers for the survey questions, and (4) use the exercise settings on the Fitbit Versa smartwatch.
- The majority of participants (ie,  $\geq 51\%$ ) will report being able to (1) read and (2) understand the daily intervention sessions on their phone.
- Participants will be given physical activity recommendations that appropriately consider their reported constraints.
- The majority of participants (ie,  $\geq 51\%$ ) who remember receiving savoring questions will report being able to understand and follow them.

#### Sustainability and feasibility:

- The auto-detection algorithm for physical activity will correctly trigger event-contingent ecological momentary assessment (EMA) prompts  $\geq 51\%$  of the time.
- Less than 25% of participants will need to be sent a new Fitbit due to technical issues.
- A total of <70% of participants will report feeling dissatisfied with the Fitbit Versa (1) exercise settings, (2) prompt notifications, and (3) prompt burden.
- In addition, <25% of participants will be sent repeated (ie, >1) reminders to open their Fitbit app and sync their study data.
- Moreover, <70% of participants who remember completing daily intervention sessions will report feeling dissatisfied with the daily intervention sessions.
- Additionally, <70% of participants will report feeling dissatisfied with the Fitbit app.
- In total, <70% of participants who remember receiving physical activity recommendations will report feeling dissatisfied with the physical activity recommendations.
- Furthermore, <70% of participants who remember receiving savoring questions will report feeling dissatisfied with the savoring questions.

#### Equity:

- Plausibility and accessibility of the eMOTION Intervention will be approximately equal between groups (ie, sex, race, ethnicity, age, BMI, income, mobility, and physical constraints).

## Methods

### Study Design

A 3-week formative study was used to test DTx RWE Framework benchmarks. Treatment components targeting intensity-based goals and affect-based goals were examined separately in different groups of participants. Stratified random assignment placed participants (N=36) into one of four groups (n=9 each): (1) intensity-based goals, (2) affect-based goals with the type and context enhancement, (3) affect-based goals with the savor enhancement, and (4) affect-based goals with type, context, and savor enhancements. Published recommendations for proof-of-concept and formative studies using similar measures to our own support the use of

homogenous samples as small as 5 to 10 subjects per user group [30-32]. Our sample (n=9 per group) precludes comparisons via inferential statistics, but corresponds with current best practices by prioritizing efficient iterations central to early-stage work—and is ultimately still amenable to descriptive statistics to determine whether the DTx meets preliminary benchmarks. Strata within each condition were based on age (ie, <45 vs  $\geq 45$  years), sex at birth (ie, male vs female), and BMI classification (ie, those with overweight vs those with obesity) to allow for approximately equal numbers of subgroup participants within each treatment group. The study was triple-blinded so that participants, research staff interacting with participants, and statisticians conducting analyses were all unaware of group

assignment. Methods were preregistered in Open Science Framework [7].

## Participants

Rolling recruitment occurred from September 2023 to March 2024 via ResearchMatch, a national health volunteer registry

### Textbox 2. Eligibility criteria.

Inclusion criteria were as follows:

- Ages 18 years and older.
- Residing in the United States.
- BMI  $\geq 25$  kg/m<sup>2</sup>.
- Current engagement in  $\leq 60$  minutes of structured physical activity (PA) per week.
- Smartphone ownership.
- Residing in an area with Wi-Fi connectivity.
- Able to speak and read English.
- Interested in starting a PA program.
- Willing to wear a smartwatch every day.
- Able to read the small font on a smartwatch screen.

Exclusion criteria screened out individuals who:

- Had cognitive disabilities precluding participation and/or the ability to provide informed consent.
- Had physical disabilities or medical issues limiting PA engagement.
- Were unable or unwilling to wear a smartwatch.
- Were currently pregnant.

## Ethical Considerations

All procedures were conducted remotely, and approval was provided by the University of Southern California Institutional Review Board (ID UP-22-00332). Eligible individuals provided written informed consent to participate in study procedures. Data were not deidentified or anonymized, so a number of protections were implemented to maintain participant privacy and confidentiality. These included the use of pseudonyms in place of names to organize study records, as well as the storage of data on secure databases and encrypted university servers. Participants were compensated up to US \$75 for completing all study procedures.

## Procedures

Following informed consent, participants completed a baseline questionnaire via REDCap (Research Electronic Data Capture; Vanderbilt University) [33,34]. Onboarding included a 1- to 3-day trial that was “passed” by completing two brief text message surveys within a single day. Participants who “passed” the trial received a Fitbit Versa 3 smartwatch in the mail and attended a 45-minute orientation session to learn about participation activities, finish technological setup, and review a PA guidebook. For the first week of the study, assessment components (eg, Fitbit smartwatch monitoring and check-ins) were introduced gradually. Then, for the following 2 weeks, participants engaged in the full DTx. Participants were asked to wear the smartwatch on their nondominant wrist for  $\geq 23$

hours/day. Smartwatch-based ecological momentary assessment (EMA) prompts (ie, “check-ins”) were delivered during PA sessions to measure affective responses to PA. EMA is a methodology involving brief repeated surveys that capture subjective experiences as they unfold in real time [35]. Participants were also instructed to press an exercise button on the smartwatch before starting a planned PA session. Researchers performed remote monitoring of participant compliance and contacted participants if missing data were detected in the study server. Finally, an exit interview and poststudy questionnaire collected feedback relating to the formative study’s SMART benchmarks.

## DTx Intervention

The eMOTION DTx was a multicomponent treatment package that was delivered under unsupervised, real-world settings and included the following elements.

### PA Guidance and Scheduling

Participants were given a prescription to start at their own pace and comfort level, slowly increasing MPA volume over the course of the study. Because PA was performed asynchronously, a digital guidebook was provided as a reference summarizing strategies to gauge PA intensity and proper warm-up and cool-down procedures. Additionally, on Sundays, participants completed a smartphone-based scheduling survey by indicating the days they intended to engage in PA for that upcoming week.

### **Daily PA Goal Sessions**

On days participants reported they intended to engage in PA, they completed 2 smartphone-based interactive intervention sessions: one in the morning and one in the evening. A text message push notification and up to 4 reminder notifications (once per hour) with a REDCap link alerted participants to complete each session. Morning sessions were scheduled for 6 AM local time. They provided a PA goal for the day (“goal module”) that differed according to the participant’s random group assignment. Participants then created a plan (“action plan module”) to achieve their goal by specifying details, including the time they planned to engage in PA that day. The final module (“coping plan module”) asked participants to anticipate barriers and brainstorm potential solutions to reach their goal. Text messages for evening sessions were sent at 7 PM local time and featured a “self-monitoring module” that asked participants to reflect on whether they met their PA goal.

In the intensity-based goal condition, participants were asked to maintain a target heart rate during PA. Target heart rate was based on the approximate age-adjusted maximum heart rate (HRmax;  $HR_{max} = 207 - [0.7 \times \text{age}]$ ), reflecting 55% HRmax for week one and 60% HRmax for week two. In the affect-based goal condition, participants were asked to engage in a type (25% of daily goals) or context (75% of daily goals; alternating places, social situations, and listening contexts) of PA that they found enjoyable. The instructions were generic to encourage self-selection (eg, “Today, your goal is to try a TYPE of PA that makes the experience more enjoyable while you are doing it”).

### **Enhancements**

Participants randomized to receive the type and context enhancement were provided tailored recommendations for PA types or contexts most likely to satisfy their unique psychological needs, as measured at baseline (ie, relatedness, social esteem, individual esteem, creativity, mindfulness, learning, challenge, entertainment, escapism, aesthetic appreciation, and morality). Recommendations were generated by an algorithm trained on data from a crowdsourced panel of adults on Amazon Mechanical Turk who were asked to rate a range of PA types and contexts on the extent to which they satisfied psychological needs. Further details on the algorithm are outside the scope of this paper. Recommendations suggested PA types or contexts matched to the participant’s most personally important psychological needs categories and eliminated types and contexts from consideration that coincided with any constraints reported at baseline.

Participants who were randomized to receive the SAVOR enhancement engaged in a brief savoring exercise to enhance the positive emotions experienced during their PA session. Immediately following the morning intervention session, these participants received a text message instructing them to click a REDCap link to complete the savoring exercise after they performed their planned PA session. The savoring exercise was comprised of two open-ended questions about the PA (eg, “Reflect on the most rewarding aspect of your PA session today. What made it feel like a success for you? How did this influence your experience?”) and surrounding environment (eg, “Let’s

take a moment to focus on the environment in which you were active today. Can you describe what you saw, felt, heard, and smelled in that space? What aspects of the environment are you grateful for?”).

## **Measures**

### **Overview of Measures**

Study variables evaluating DTx performance for SMART benchmarks were collected via questionnaires, videoconference exit interviews, and EMA. During exit interviews, if the participant’s initial response to an open-ended question was too vague, the surveyor followed up with neutrally worded clarifying questions. During response coding for analysis, if any response contained conflicting statements and true clarity could not be achieved, the response was always treated conservatively (ie, biased toward the null hypothesis). Measures are briefly summarized by benchmark type below and are elaborated in detail in [Multimedia Appendix 1](#).

### **Safety**

Safety was assessed by calculating the number of adverse events reported via official Institutional Review Board (IRB) channels, which allow participants and study teams to report events. Participants were given IRB contact information for this purpose during informed consent.

### **Plausibility**

Plausibility was primarily assessed with the following exit interview question: “When comparing your experience with PA before and after participation in the eMOTION Study, do you personally feel as if the degree of enjoyment you feel while engaging in PA changed or stayed the same?” Responses were transcribed, and enjoyment was coded as “decreased” (0), “stayed the same” (1), or “increased” (2).

### **Accessibility and Usability**

#### **Daily Intervention Sessions and Fitbit Smartwatch**

The poststudy questionnaire included the 10-item System Usability Scale (SUS) to assess the accessibility and usability of the daily intervention sessions and the Fitbit smartwatch [36]. Participants indicated the extent to which they agreed with each statement using a Likert Scale (“strongly disagree” {0} to “strongly agree” {4}). Negative items were reverse-scored, all items summed, and the sum multiplied by 2.5 to yield a total score ranging from 0 to 100. Items displayed acceptable reliability (Cronbach  $\alpha=0.78$  and  $0.91$  for daily goals and Fitbit smartwatch, respectively). Participants were also asked investigator-created questions in the exit interview about their ability to read and understand the wording of the daily intervention sessions and check-ins; select their answers for the check-ins; and use the exercise buttons on the smartwatch. Responses were transcribed and coded as “unable” (0) or “able” (1).

#### **Tailored Physical Activity Recommendation Enhancement**

Multiple measures assessed the accessibility and usability of type and context recommendations. During the exit interview, participants were asked whether they were “generally able to follow the PA [type and context] recommendations” they

received. Responses were transcribed and coded as “unable” (0) or “able” (1). Reported inability was then considered alongside constraints reported at baseline. Specifically, in the baseline questionnaire, a set of questions asked about constraints related to PA type, location, and listening contexts. Participants indicated their ability to engage in 47 different activity types [37] by selecting as many options as applied for each one: “able to do it,” “physically unable to do it,” “no place to do it,” “no equipment to do it,” and “I don’t know what this is.” For some activities, “no equipment to do it” was not considered a constraint because no equipment was required (eg, bodyweight exercises); for others, “I don’t know what this is” did not apply because participants were provided with instructional videos for the activity. Next, participants indicated their ability to access 19 different locations and facilities for PA by selecting “I’m able” or “I don’t have access or can’t get there” for each. Finally, participants selected “yes” or “no” for whether they were able to listen to music, audiobooks, or other media on their smartphone during PA. The total number of possible constraints across PA types and contexts ranged from 0 to 68.

### Savoring Enhancement

Participants were asked during the exit interview whether they were able to understand and follow the savoring exercises. Responses were transcribed and coded as “unable” (0) or “able” (1).

### Sustainability and Feasibility

#### Satisfaction With DTx Components

Most benchmarks for sustainability and feasibility were assessed with the Delighted-Terrible Scale [38,39] in the poststudy questionnaire. Participants were asked to indicate how they felt about the smartwatch exercise settings; check-in notifications and burden; daily intervention sessions; smartphone data syncing procedures; and type, context, and/or savor enhancements, as applicable. Multiple choice options for each item were: “delighted” (8), “pleased” (7), “mostly satisfied” (6), “mixed-about equally satisfied and dissatisfied” (5), “mostly dissatisfied” (4), “unhappy” (3), “terrible” (2), “neutral- neither satisfied nor dissatisfied” (1), and “I never thought about it” (0). Response options “mostly dissatisfied,” “unhappy,” and “terrible” were considered as dissatisfaction with the given DTx component.

#### Fidelity of PA Auto-Detection Algorithm

An algorithm triggered check-ins when the smartwatch sensors detected PA engagement via sufficiently elevated 10-minute rolling average heart rate (ie, 55% to 60% of age-adjusted HR<sub>max</sub>). Heart rate data from smartwatch sensors were considered alongside Fitabase software prompting records, which contained timestamps associated with the timing of smartwatch check-ins.

#### Research Staff Burden

Some measures assessed whether DTx-associated workloads were sustainable and feasible for research staff to maintain. A REDCap form was used to track smartwatch inventory and document technical difficulties requiring device replacement. Staff also recorded all instances where they had to remind

participants to sync their smartphone data due to >48 hours elapsing without data reaching Fitabase servers.

### Equity

To explore equity, the proportions of participants recalling increased PA enjoyment after the study, and rating the daily goals and smartwatch at or above the acceptable cutoff point (ie, SUS score  $\geq 68.0$ ), were compared across sex, race, ethnicity, age, BMI, income, mobility, and physical constraints subgroups (subjective comparisons, given a lack of statistical power). Subgroup data were collected with screening and baseline questionnaires. Sex (male or female), race (White or non-White), ethnicity (Hispanic or non-Hispanic), and income (unsure;  $\leq$ US \$44,999; US \$45,000-\$84,999; US \$85,000-\$124,999; or  $\geq$ US \$125,000) were directly assessed with single items. Age was calculated as the difference between date-of-birth and baseline questionnaire completion date, then split into quartiles ( $\leq 37$  y, 38-43 y, 44-58 y, and  $\geq 59$  y). BMI was calculated using reported height and weight and coded according to established clinical cutoff points (those with overweight  $\leq 29.9$ , those with obesity  $\geq 30.0$ ). Mobility was assessed with an investigator-developed measure, which asked participants to indicate the degree to which they agreed with 5 different statements (eg, “I have had trouble bending over”) using a 10-point sliding scale with the markers “not at all true” (1), “moderately true” (5), and “completely true” (10). Items were summed into a total mobility score (range 5 [least trouble] to 50 [most trouble]), which was then split into quartiles ( $\leq 13.5$ , 13.6-18.5, 18.6-28.5, and  $\geq 28.6$ ). Finally, for physical constraints, activity types designated as “physically unable to do it” by the participant at baseline were tallied to create a total number ranging from 0 to 47, which was split into tertiles (0, 1-4;  $\geq 5$ ).

### Statistical Analysis

The formative study was not statistically powered to test hypotheses using inferential statistics. Rather, as is common for proof-of-concept trials, the sample size was chosen to balance monetary and personnel constraints while allowing the DTx to be applied and assessed in the target population. To minimize confirmation bias and reduce the risk of moving forward with a large-scale phase III comparison study prematurely, descriptive statistics were used to compare formative study evaluation data against a priori benchmarks with clear go or no-go criteria [8]. Sensitivity analyses featuring linear and logistic regression models were performed to compare person-level characteristics between the final sample and participants lost to follow-up. All analyses were performed using IBM SPSS Statistics software (version 29).

## Results

### Sample Characteristics

A total of 46 participants consented and enrolled in the eMOTION Formative Study. Of these, 9 participants left the study during onboarding and before the intervention: 2 withdrew due to personal obligations (ie, surgery and work demands), 1 was unenrolled due to technical difficulties associated with responding to text message EMA on their personal smartphone,

and 6 were lost to follow-up without explanation (ie, study staff were unable to reestablish contact).

**Table 1** provides baseline characteristics for the 37 participants who completed the intervention. This baseline sample was 46% female (n=17), 86% non-Hispanic (n=32), and 86% White (n=32). Mean age was 46.2 (SD 13.4) years. A broad range of household incomes was represented (**Table 1**). One additional participant in the sample was lost to follow-up after completing the intervention but before the poststudy questionnaire and exit

interview. Therefore, a final analytic sample of 36 participants who completed the study and provided at least partial outcome data was retained. Sensitivity analyses confirmed that the final sample did not differ from participants lost to follow-up on the basis of their sex ( $P=.88$ ), age ( $P=.82$ ), or BMI ( $P=.98$ ).

The baseline sample comprised 37 participants, while 36 participants provided at least some information for the main outcomes of interest by filling out the poststudy questionnaire and/or completing the exit interview.

**Table 1.** Participant characteristics assessed at baseline.

Characteristics	Values
Age (years), mean (SD)	46.2 (13.4)
<b>Sex, n (%)</b>	
Male	17 (46)
Female	20 (54)
<b>Ethnicity, n (%)</b>	
Non-Hispanic	32 (86)
Hispanic	5 (14)
<b>Race, n (%)</b>	
White	32 (86)
Black	4 (11)
Asian	1 (3)
Native American	0 (0)
Pacific Islander	0 (0)
Multiracial	0 (0)
<b>Household income (US \$), n (%)</b>	
≤44,999	7 (19)
45,000 to 84,999	9 (24)
85,000 to 124,999	8 (22)
≥125,000	10 (27)
Unsure	3 (8)

## SMART Benchmarks

### Safety

No adverse events were reported during the study. With 0% of participants experiencing an adverse event, the safety benchmark (<1% of participants) was met.

### Plausibility

Of the 27 participants who received affect-based goals, 14 (52%) reported during the exit interview that PA enjoyment increased over the course of the study (ie, comparing enjoyment before the study vs after the study), while 13 (48%) said enjoyment stayed the same. Therefore, the plausibility benchmark was just met (ie, ≥51% recalled an overall increase in enjoyment).

## Accessibility and Usability

### Daily Intervention Sessions and Fitbit Smartwatch

The daily intervention sessions and Fitbit smartwatch received average total SUS scores of 72.4 (SD 17.2) and 73.1 (SD 22.9), respectively. These scores exceeded the prespecified cutoff point in the SMART benchmarks (ie, ≥68). Post hoc calculations additionally determined that 23 (64%) and 26 (72%) participants rated the daily intervention sessions and Fitbit smartwatch components at-or-above the cutoff point (ie, ≥68). The majority of participants also reported being able to read the smartwatch check-ins (n=34, 97%) and daily intervention sessions (n=34, 97%), understand the check-ins (n=32, 91%) and daily sessions (n=34, 97%), select answers for check-ins (n=33, 94%), and use exercise settings on the smartwatch (n=27, 77%), therefore meeting the associated benchmarks.

### Type and Context Enhancement

Although 18 participants received type and context recommendations, 3 did not remember receiving them. Table 2 depicts results from the 15 participants who received and remembered type and context recommendations. There was no discernable relationship between the number of activity constraints and reported ability to follow type and context recommendations, suggesting that the benchmark was met (ie,

participants were given PA recommendations that appropriately considered their constraints).

Only data from participants who both received and remembered receiving PA type and context recommendations (ie, n=15) were included in the table. An additional 3 participants received, but did not remember receiving, recommendations; 5 participants incorrectly reported that they received type and context (ie, when they did not receive type and context); these data were excluded from the table.

**Table 2.** Accessibility of type and context recommendations.

PA <sup>a</sup> constraint quartile	Reported being able to follow recommendations, n/N (%)
1	2/5 (40)
2	1/2 (50)
3	2/4 (50)
4	1/4 (25)

<sup>a</sup>PA: physical activity.

### SAVOR Enhancement

Although 18 participants received savoring exercises, 4 did not remember receiving them. Of the 14 participants who received and remembered SAVOR, 12 (86%) reported being able to understand and follow the savoring exercises. Therefore, the associated benchmark ( $\geq 51\%$  of participants who remember receiving savoring exercises will report being able to understand and follow them) was met.

### Sustainability and Feasibility

#### Satisfaction With DTx Components

The proportion of participants reporting dissatisfaction with a DTx component (ie, Fitbit smartwatch exercise settings, check-in notifications, and check-in burden; daily intervention sessions; Fitbit app; PA type and context recommendations; and savoring exercises) ranged from 0% (n=0) to 39% (n=14; Table 3). Therefore, the associated benchmarks were met.

**Table 3.** Proportion of participants reporting dissatisfaction with digital therapeutics components.

Benchmark	Result, n (%) <sup>a</sup>
<b>&lt;70% of participants will report feeling dissatisfied with Fitbit features</b>	
Exercise settings	6 (17)
Check-in notifications	12 (33)
Check-in burden	14 (39)
<70% of participants will report feeling dissatisfied with the Fitbit app	3 (8)
<b>&lt;70% of participants who remember an intervention component will report feeling dissatisfied with it</b>	
Daily intervention sessions <sup>b</sup>	3 (9)
PA recommendations <sup>c</sup>	3 (23)
Savoring exercises <sup>d</sup>	0 (0)

<sup>a</sup>N=36 unless otherwise indicated.

<sup>b</sup>n=34 (2 participants did not remember receiving daily intervention sessions).

<sup>c</sup>n=13 (15 remembered receiving type and context recommendations, but 2 did not respond to the satisfaction item).

<sup>d</sup>n=12 (14 remembered receiving SAVOR, but 2 did not respond to the satisfaction item).

### Fidelity of PA Auto-Detection Algorithm

All smartwatch check-ins were correctly sent when the rolling 10-minute average heart rate crossed the threshold specified in the algorithm (ie, 55%-60% HRmax). Therefore, the triggering algorithm achieved 100% accuracy and met the threshold (ie,

the algorithm triggered smartwatch check-ins correctly  $\geq 51\%$  of the time).

### Research Staff Burden

No replacement Fitbit smartwatches were sent out to participants due to technical issues; therefore, the meeting did not meet the associated benchmark (ie, <25% of participants needed a new

Fitbit due to technical issues). Additionally, only 6/36 (ie, 17%) participants were sent a single reminder to sync their Fitbit data to their smartphone, and no participants (ie, 0%) required more than one reminder. Therefore, the benchmark (<25% of participants will be sent repeated [ $>1$ ] reminders to open their Fitbit app and sync their study data) was met.

### Equity

Results for the equity benchmark are summarized in [Table 4](#). Values are the proportion of participants in each subgroup who reported increased PA enjoyment and who scored the daily goals and Fitbit smartwatch at or above the accessibility and usability cutoff point. Cell sizes were too small to use chi-square tests between groups. Subjective comparisons generally suggested equal proportions across group levels, with a few notable exceptions. First, plausibility differed by sex, with 64% (n=9) of women reporting increased enjoyment during the study compared to only 38% (n=5) of men. Additionally, non-White participants reported greater accessibility, usability (4/5 to 5/5, 80%-100% vs 19/31 to 21/31, 61%-68%), and plausibility (3/4, 75% vs 11/23, 48%) compared to White participants.

Accessibility and usability were comparable across all income categories except for the highest income group, 90% (9/10) of whom rated the daily goals and Fitbit smartwatch at-or-above the cutoff point (compared to 3/6 and 4/8, 50% to 6/9, 69% in the other groups). Finally, accessibility, usability, and plausibility were lowest among individuals reporting the most physical constraints compared to those reporting fewer than 5 constraints or no constraints, as well as those reporting greater mobility impairment.

In [Table 4](#), proportions and percentages for accessibility and usability reflect participants who scored the daily intervention sessions and Fitbit smartwatch, respectively, at-or-above the System Usability Scale cutoff of 68. Proportions and percentages in the plausibility column are the participants in the affect-based groups who reported increased PA enjoyment over the course of the study. A total of 36 completed the accessibility and usability items in the poststudy questionnaire, and 35 completed the plausibility item in the exit interview. Cell sizes were too small to conduct chi-Square comparisons, so potential group differences were subjectively assessed.

**Table 4.** Equity benchmark performance.

Characteristics	Accessibility and usability		Plausibility, n/N (%)
	Daily goals, n/N (%)	Fitbit, n/N (%)	
<b>Sex</b>			
Male	11/17 (65)	13/17 (76)	5/13 (38)
Female	12/19 (63)	13/19 (68)	9/14 (64)
<b>Ethnicity</b>			
Non-Hispanic	21/32 (66)	24/32 (75)	13/25 (52)
Hispanic	2/4 (50)	2/4 (50)	1/2 (50)
<b>Race</b>			
White	19/31 (61)	21/31 (68)	11/23 (48)
Non-White	4/5 (80)	5/5 (100)	3/4 (75)
<b>Household income (US \$)</b>			
≤44,999	3/6 (50)	3/6 (50)	1/5 (20)
45,000 to 84,999	6/9 (67)	6/9 (67)	7/7 (100)
85,000 to 124,999	4/8 (50)	5/8 (63)	3/7 (43)
≥125,000	9/10 (90)	9/10 (90)	3/5 (60)
Unsure	1/3 (33)	3/3 (100)	0/3 (0)
<b>Age (years)</b>			
≤37	3/9 (33)	6/9 (67)	4/7 (57)
38-43	6/9 (67)	6/9 (67)	2/7 (29)
44-58	8/9 (89)	8/9 (89)	4/5 (80)
≥59	6/9 (67)	6/9 (67)	4/8 (50)
<b>BMI (kg/m<sup>2</sup>)</b>			
Overweight	8/12 (67)	10/12 (83)	5/10 (50)
Obese	15/24 (63)	16/24 (67)	9/17 (53)
<b>Mobility score</b>			
≤13.5	6/9 (67)	7/9 (78)	5/7 (71)
13.6-18.5	6/9 (67)	8/9 (89)	3/6 (50)
18.6-28.5	6/9 (67)	7/9 (78)	4/8 (50)
≥28.6	5/9 (56)	4/9 (44)	2/6 (33)
<b>Physical constraints</b>			
No constraints	10/15 (67)	12/15 (80)	5/11 (45)
1-5 constraints	9/12 (75)	9/12 (75)	6/8 (75)
≥5 constraints	4/9 (44)	5/9 (56)	3/8 (38)

## Discussion

### Principal Findings

The novel eMOTION DTx was systematically designed, developed, and formatively tested according to benchmarks recommended by the DTx RWE Framework. Multiple evidence-based intervention components were integrated to address the problem of inactivity due to low PA enjoyment among adults with overweight or obesity. Daily digitally interactive intervention sessions featured Health Action Process Approach self-regulatory strategies (ie, action planning, coping

planning, and self-monitoring) to encourage attainment of PA goals [28]. Affect and Health Behavior Framework mechanisms, including affective responses during PA, were targeted with affect-based PA goals, as well as through two intervention enhancements [40]. The type and context enhancement involved a tailoring algorithm that generated recommendations for specific types and contexts of PA likely to satisfy personally relevant psychological needs. This was informed by a body of evidence demonstrating that the degree of importance ascribed to higher-order psychological needs varies between-persons [41-44], as well as by our preliminary work suggesting that

various PA types and contexts differentially satisfy psychological needs and enhance positive affect during PA [45,46]. The SAVOR enhancement leveraged savoring exercises to intensify the influence of affective responses on PA [47,48]. Savoring is a mindfulness practice that enhances the saliency of positive experiences via attentional deployment and gratitude-making [49,50]. The DTx underwent rigorous internal testing and iterative refinements during phase I to arrive at a minimum viable product for application in phase II.

## Principal Results

SMART benchmarks determining success were defined a priori to maximize transparency and preregistered in accordance with open science best practices. Final benchmark results fully supported the safety and plausibility of the eMOTION DTx. Standards for safety and plausibility are considered the most important to uphold, with the Food and Drug Administration's various approval routes all requiring new medical devices to demonstrate sufficient safety and efficacy before public release.

Findings also confirmed eMOTION DTx accessibility and usability. Accessibility and usability were assessed for multidevice (ie, smartphone and Fitbit smartwatch) user interface, study-related functions, and specific intervention content based on previous work, underscoring the importance of these areas for participant engagement and intervention delivery [51-53]. While objective sustainability and feasibility benchmarks were all met, minor real-time iterative adjustments were made to decrease smartwatch check-in burden in response to participant feedback. Early into formative testing, many participants reported receiving a high volume of check-ins throughout the day that were only meant to be received during PA sessions. To increase the specificity of the algorithm (ie, improve differentiation between low-intensity lifestyle activities and structured MPA sessions), the original heart rate threshold (ie, 10-minute rolling average  $\geq 50\%$  HRmax) was increased to 55% HRmax.

In an effort to increase the validity of accessibility, usability, sustainability, and feasibility feedback, participants were asked whether they remembered receiving each of the intervention components (ie, daily intervention sessions, PA recommendations, and savoring exercises). Terminology for these questions was consistent with that used during onboarding and throughout all study materials. Only those who received and remembered receiving a given component were included in the respective analysis. Interestingly, we found that some participants had forgotten the intervention components they received during the study (ie, 2 did not remember receiving daily goal intervention sessions; 3, the tailored PA recommendations; and 4, the savoring exercises). This same general phenomenon has been described by other researchers engaging in complex, multicomponent interventions, in which participants fail to retrospectively identify—or otherwise misremember or misunderstand—critical aspects of the study [54,55]. By comparing objective study records with participant self-reports, we were able to control for this form of information bias [56].

Finally, equity in accessibility, usability, and plausibility was generally observed (via subjective comparisons) across sex,

race, ethnicity, income, age, BMI, mobility, and physical constraints subgroups, although notable discrepancies did arise. First, a greater proportion of women reported increased PA enjoyment during the study compared to men. Previous work presents conflicting evidence for whether men or women are more likely to enjoy PA [57,58]. However, our findings might reflect those of others that consistently suggest the presence of sex-based differences in the relevance and plausibility of mHealth interventions for PA. Specifically, women are more likely to use PA-tracking mHealth apps [59], report using PA apps for the purpose of enjoyment [60], and experience more positive health outcomes following PA mHealth interventions compared to men [61]. DTx accessibility, usability, and plausibility also differed by race, but—contrary to what might be expected—favored non-White participants. This result is likely unreliable due to imbalanced cell sizes (ie, 31 White and 5 non-White) and was not necessarily a cause for concern, as the DTx was more accessible, usable, and efficacious for the traditionally disadvantaged group. Racial differences will be further explored using a larger sample in Phase III. More concerning trends were observed for household income, mobility, and physical constraint subgroups. Specifically, the DTx was most accessible and usable for the wealthiest and least efficacious for the lowest-income participants. The DTx was also the least accessible, usable, and least efficacious for participants with the most physical constraints and most limited mobility, respectively. To address these findings, a free YouTube channel was created for participants. The channel included exercise videos using positive, upbeat formats and featuring diverse instructors, contexts, and PA types. Videos required only minimal space with no equipment necessary, and provided optional physical adaptations for each exercise. These changes sought to provide a larger number of enjoyable PA options that were convenient, accessible, and affordable to all participants, regardless of their income or physical abilities.

## Limitations

The extant literature contained very few validated measures for smartphone DTx accessibility, usability, sustainability, and feasibility. Therefore, we adapted existing measures from related technologies (eg, online websites and software systems) and created our own when appropriate. Investigator-developed items were highly specific to our particular DTx, potentially limiting generalizability to other DTx setups. Additionally, this phase II study's small sample size precluded the use of inferential statistics; findings were descriptive, therefore, and should be interpreted with caution. Similarly, results for safety, plausibility, accessibility, usability, sustainability, feasibility, and equity of the eMOTION DTx reflect preliminary findings in our small, homogenous sample. Specific subgroups were at times collapsed into more general “affect-based” and “intensity-based” groups, and we were unable to explore whether specific combinations of enhancements drove the main effects. However, the phase III study will feature a sufficiently powered, diverse sample (N=280) to rigorously analyze within- and between-group differences in DTx efficacy and optimize treatment components.

## Conclusion

Overall, this study is one of the first to apply benchmarks recommended by the DTx RWE Framework to formatively test a novel digital intervention. By transparently reporting the specific measurement and analytic choices we made to operationalize the DTx RWE Framework, our study hopes to contribute methodology and results that can be replicated. This is an important undertaking, since recent investigations suggest that the inconsistent application of frameworks is hindering the advancement of knowledge in real-world implementation contexts [62]. Future work should aim to apply the DTx RWE Framework in other contexts to determine whether its processes are relevant across diverse health problems, target populations, and DTx formats.

All benchmarks were determined to be met satisfactorily following iterative improvements made to the eMOTION DTx during development. Preliminary evidence, therefore, supports a research transition into DTx RWE Framework phase III, in which a large-scale effectiveness trial will determine whether the eMOTION DTx produces a greater effect (ie, PA enjoyment) among the experimental (compared to control) group. Ultimately, our application of the DTx RWE Framework ensures that the design and development of the eMOTION DTx follow current best practices, increase the likelihood that regulatory standards are met, and facilitate an efficient translation from research to real-world practice.

## Acknowledgments

This study was funded by a grant from the National Cancer Institute (R01CA272933). RC-L's effort on this paper was supported by a grant from the National Heart, Lung, and Blood Institute (F31HL176165). The authors attest that no generative artificial intelligence technology was used to generate text, figures, or other informational content in this manuscript.

## Funding

This study was funded by a grant from the National Cancer Institute (R01CA272933). Rachel Crosley-Lyons' effort on the current paper was supported by a grant from the National Heart Lung and Blood Institute (F31HL176165).

## Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

## Authors' Contributions

Conceptualization: RC-L, EH, GFD Methodology: RC-L, EH, GFD Formal analysis: RC-L, WLW Investigation: RC-L, LH, MH, GFD Data curation: RC-L, LH, WLW Software: LH, WLW Resources: GFD Writing – original draft: RC-L, MH, EH Writing – review & editing: RC-L, LH, MH, WLW, DVD, JH, EH, GFD Visualization: RC-L Project administration: MH Supervision: DVD, JH, EH, GFD Funding acquisition: GFD

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Benchmark measurement and assessment details.

[\[DOCX File, 50 KB-Multimedia Appendix 1\]](#)

## References

1. Hacker K. The burden of chronic disease. *Mayo Clin Proc Innov Qual Outcomes*. 2024;8(1):112-119. [FREE Full text] [doi: [10.1016/j.mayocpiqo.2023.08.005](https://doi.org/10.1016/j.mayocpiqo.2023.08.005)] [Medline: [38304166](https://pubmed.ncbi.nlm.nih.gov/38304166/)]
2. Grand View. Digital Therapeutics Market Size And Share Report, 2030. Grand View Research; 2024. URL: <https://tinyurl.com/2ttfdrte> [accessed 2024-09-09]
3. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: The ORBIT model for developing behavioral treatments for chronic diseases. *Health Psychol*. 2015;34(10):971-982. [FREE Full text] [doi: [10.1037/hea0000161](https://doi.org/10.1037/hea0000161)] [Medline: [25642841](https://pubmed.ncbi.nlm.nih.gov/25642841/)]
4. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124. [FREE Full text] [doi: [10.1371/journal.pmed.0020124](https://doi.org/10.1371/journal.pmed.0020124)] [Medline: [16060722](https://pubmed.ncbi.nlm.nih.gov/16060722/)]
5. Ioannidis JPA. Why most discovered true associations are inflated. *Epidemiology*. 2008;19(5):640-648. [doi: [10.1097/EDE.0b013e31818131e7](https://doi.org/10.1097/EDE.0b013e31818131e7)] [Medline: [18633328](https://pubmed.ncbi.nlm.nih.gov/18633328/)]
6. Man-Son-Hing M, Laupacis A, O'Rourke K, Molnar FJ, Mahon J, Chan KBY, et al. Determination of the clinical importance of study results. *J Gen Intern Med*. 2002;17(6):469-476. [FREE Full text] [doi: [10.1046/j.1525-1497.2002.11111.x](https://doi.org/10.1046/j.1525-1497.2002.11111.x)] [Medline: [12133163](https://pubmed.ncbi.nlm.nih.gov/12133163/)]

7. Crosley-Lyons R, Dunton G, Hewus M, Hatzinger L, Wang W. Using Real-Time Data Capture to Examine Affective Mechanisms as Mediators of Physical Activity Adherence in Interventions: The eMOTION Formative Study. Published online April 17, 2024. URL: <https://osf.io/QTF79> [accessed 2025-12-01]
8. Kim M, Patrick K, Nebeker C, Godino J, Stein S, Klasnja P, et al. The digital therapeutics real-world evidence framework: an approach for guiding evidence-based digital therapeutics design, development, testing, and monitoring. *J Med Internet Res*. 2024;26:e49208. [FREE Full text] [doi: [10.2196/49208](https://doi.org/10.2196/49208)] [Medline: [38441954](https://pubmed.ncbi.nlm.nih.gov/38441954/)]
9. Dang A, Arora D, Rane P. Role of digital therapeutics and the changing future of healthcare. *J Family Med Prim Care*. 2020;9(5):2207-2213. [FREE Full text] [doi: [10.4103/jfmpe.jfmpe\\_105\\_20](https://doi.org/10.4103/jfmpe.jfmpe_105_20)] [Medline: [32754475](https://pubmed.ncbi.nlm.nih.gov/32754475/)]
10. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *Br J Sports Med*. 2020;54(24):1451-1462. [FREE Full text] [doi: [10.1136/bjsports-2020-102955](https://doi.org/10.1136/bjsports-2020-102955)] [Medline: [33239350](https://pubmed.ncbi.nlm.nih.gov/33239350/)]
11. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*. 2008;40(1):181-188. [doi: [10.1249/mss.0b013e31815a51b3](https://doi.org/10.1249/mss.0b013e31815a51b3)] [Medline: [18091006](https://pubmed.ncbi.nlm.nih.gov/18091006/)]
12. McEachan RRC, Conner M, Taylor NJ, Lawton RJ. Prospective prediction of health-related behaviours with the Theory of Planned Behaviour: a meta-analysis. *Health Psychology Review*. 2011;5(2):97-144. [doi: [10.1080/17437199.2010.521684](https://doi.org/10.1080/17437199.2010.521684)]
13. Turner RR, Steed L, Quirk H, Greasley RU, Saxton JM, Taylor SJ, et al. Interventions for promoting habitual exercise in people living with and beyond cancer. *Cochrane Database Syst Rev*. 2018;9(9):CD010192. [FREE Full text] [doi: [10.1002/14651858.CD010192.pub3](https://doi.org/10.1002/14651858.CD010192.pub3)] [Medline: [30229557](https://pubmed.ncbi.nlm.nih.gov/30229557/)]
14. Young MD, Plotnikoff RC, Collins CE, Callister R, Morgan PJ. Social cognitive theory and physical activity: a systematic review and meta-analysis. *Obes Rev*. 2014;15(12):983-995. [doi: [10.1111/obr.12225](https://doi.org/10.1111/obr.12225)] [Medline: [25428600](https://pubmed.ncbi.nlm.nih.gov/25428600/)]
15. Rebar AL, Rhodes RE, Gardner B. How we are misinterpreting physical activity intention - behavior relations and what to do about it. *Int J Behav Nutr Phys Act*. 2019;16(1):71. [FREE Full text] [doi: [10.1186/s12966-019-0829-y](https://doi.org/10.1186/s12966-019-0829-y)] [Medline: [31438956](https://pubmed.ncbi.nlm.nih.gov/31438956/)]
16. Rhodes R, Cox A, Sayar R. What predicts the physical activity intention-behavior gap? a systematic review. *Ann Behav Med*. 2022;56(1):1-20. [doi: [10.1093/abm/kaab044](https://doi.org/10.1093/abm/kaab044)] [Medline: [34231844](https://pubmed.ncbi.nlm.nih.gov/34231844/)]
17. Murphy S, Eaves D. Exercising for the pleasure and for the pain of it: the implications of different forms of hedonistic thinking in theories of physical activity behavior. *Front Psychol*. 2016;7:843. [FREE Full text] [doi: [10.3389/fpsyg.2016.00843](https://doi.org/10.3389/fpsyg.2016.00843)] [Medline: [27375524](https://pubmed.ncbi.nlm.nih.gov/27375524/)]
18. Stults-Kolehmainen MA, Blacutt M, Bartholomew JB, Gilson TA, Ash GI, McKee PC, et al. Motivation states for physical activity and sedentary behavior: desire, urge, wanting, and craving. *Front Psychol*. 2020;11:568390. [FREE Full text] [doi: [10.3389/fpsyg.2020.568390](https://doi.org/10.3389/fpsyg.2020.568390)] [Medline: [33240154](https://pubmed.ncbi.nlm.nih.gov/33240154/)]
19. Ekkekakis P, Lind E. Exercise does not feel the same when you are overweight: the impact of self-selected and imposed intensity on affect and exertion. *Int J Obes (Lond)*. 2006;30(4):652-660. [doi: [10.1038/sj.ijo.0803052](https://doi.org/10.1038/sj.ijo.0803052)] [Medline: [16130028](https://pubmed.ncbi.nlm.nih.gov/16130028/)]
20. Liao Y, Chou CP, Huh J, Leventhal A, Dunton G. Examining acute bi-directional relationships between affect, physical feeling states, and physical activity in free-living situations using electronic ecological momentary assessment. *J Behav Med*. 2017;40(3):445-457. [FREE Full text] [doi: [10.1007/s10865-016-9808-9](https://doi.org/10.1007/s10865-016-9808-9)] [Medline: [27766481](https://pubmed.ncbi.nlm.nih.gov/27766481/)]
21. Ekkekakis P, Parfitt G, Petruzzello SJ. The pleasure and displeasure people feel when they exercise at different intensities: decennial update and progress towards a tripartite rationale for exercise intensity prescription. *Sports Med*. 2011;41(8):641-671. [doi: [10.2165/11590680-000000000-00000](https://doi.org/10.2165/11590680-000000000-00000)] [Medline: [21780850](https://pubmed.ncbi.nlm.nih.gov/21780850/)]
22. Bartholomew JB, Miller BM. Affective responses to an aerobic dance class: the impact of perceived performance. *Res Q Exerc Sport*. 2002;73(3):301-309. [doi: [10.1080/02701367.2002.10609024](https://doi.org/10.1080/02701367.2002.10609024)] [Medline: [12230337](https://pubmed.ncbi.nlm.nih.gov/12230337/)]
23. Carraro A, Paoli A, Gobbi E. Affective response to acute resistance exercise: a comparison among machines and free weights. *Sport Sci Health*. 2018;14(2):283-288. [doi: [10.1007/s11332-018-0427-4](https://doi.org/10.1007/s11332-018-0427-4)]
24. Dunton GF, Liao Y, Intille S, Huh J, Leventhal A. Momentary assessment of contextual influences on affective response during physical activity. *Health Psychol*. 2015;34(12):1145-1153. [FREE Full text] [doi: [10.1037/hea0000223](https://doi.org/10.1037/hea0000223)] [Medline: [26053885](https://pubmed.ncbi.nlm.nih.gov/26053885/)]
25. Lacharité-Lemieux M, Brunelle JP, Dionne IJ. Adherence to exercise and affective responses: comparison between outdoor and indoor training. *Menopause*. 2015;22(7):731-740. [doi: [10.1097/GME.0000000000000366](https://doi.org/10.1097/GME.0000000000000366)] [Medline: [25423324](https://pubmed.ncbi.nlm.nih.gov/25423324/)]
26. Crosley-Lyons R, Do B, Hewus M, Dunton GF. An ecological momentary assessment study of affectively-charged motivational states and physical activity. *Psychol Sport Exerc*. 2023;67:102423. [doi: [10.1016/j.psychsport.2023.102423](https://doi.org/10.1016/j.psychsport.2023.102423)] [Medline: [37665876](https://pubmed.ncbi.nlm.nih.gov/37665876/)]
27. Dunton GF, Crosley-Lyons R, Rhodes RE. Affective response during real-world physical activity as an intervention mediator. *Exerc Sport Sci Rev*. 2023;51(4):140-149. [doi: [10.1249/JES.0000000000000321](https://doi.org/10.1249/JES.0000000000000321)] [Medline: [37195062](https://pubmed.ncbi.nlm.nih.gov/37195062/)]
28. Degroote L, Paepe AD, Bourdeaudhuij ID, Dyck DV, Crombez G. Effectiveness of the mHealth intervention 'MyDayPlan' to increase physical activity: an aggregated single case approach. *Int J Behav Nutr Phys Act*. 2021;18(1):92. [FREE Full text] [doi: [10.1186/s12966-021-01163-2](https://doi.org/10.1186/s12966-021-01163-2)] [Medline: [34233718](https://pubmed.ncbi.nlm.nih.gov/34233718/)]
29. Schwarzer R, Luszczynska A. How to overcome health-compromising behaviors. *European Psychologist*. 2008;13(2):141-151. [doi: [10.1027/1016-9040.13.2.141](https://doi.org/10.1027/1016-9040.13.2.141)]

30. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Soc Sci Med.* 2022;292:114523. [FREE Full text] [doi: [10.1016/j.socscimed.2021.114523](https://doi.org/10.1016/j.socscimed.2021.114523)] [Medline: [34785096](https://pubmed.ncbi.nlm.nih.gov/34785096/)]
31. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. *Behav Res Methods Instrum Comput.* 2003;35(3):379-383. [doi: [10.3758/bf03195514](https://doi.org/10.3758/bf03195514)] [Medline: [14587545](https://pubmed.ncbi.nlm.nih.gov/14587545/)]
32. McCombie C, Esponda GM, Schmidt U, Lawrence V. Qualitative diary methods in mental health research: A scoping review and recommendations for research and reporting. *European Psychologist.* 2024;29(1):3-16. [doi: [10.1177/02537176251363817](https://doi.org/10.1177/02537176251363817)] [Medline: [40799658](https://pubmed.ncbi.nlm.nih.gov/40799658/)]
33. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42(2):377-381. [FREE Full text] [doi: [10.1016/j.jbi.2008.08.010](https://doi.org/10.1016/j.jbi.2008.08.010)] [Medline: [18929686](https://pubmed.ncbi.nlm.nih.gov/18929686/)]
34. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform.* Jul 2019;95:103208. [FREE Full text] [doi: [10.1016/j.jbi.2019.103208](https://doi.org/10.1016/j.jbi.2019.103208)] [Medline: [31078660](https://pubmed.ncbi.nlm.nih.gov/31078660/)]
35. Stone A, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. *Ann Behav Med.* 1994;16(3):199-202. [doi: [10.1093/abm/16.3.199](https://doi.org/10.1093/abm/16.3.199)]
36. Brooke J. SUS: A Quick and Dirty Usability Scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. *Usability Evaluation in Industry.* 1st ed. Boca Raton, Florida, USA. CRC Press; 1996:189-194.
37. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Tudor-Locke C, et al. 2011 compendium of physical activities: a second update of codes and MET values. *Med Sci Sports Exerc.* 2011;43(8):1575-1581. [doi: [10.1249/MSS.0b013e31821ece12](https://doi.org/10.1249/MSS.0b013e31821ece12)] [Medline: [21681120](https://pubmed.ncbi.nlm.nih.gov/21681120/)]
38. Westbrook RA. A rating scale for measuring product/service satisfaction. *Journal of Marketing.* 1980;44(4):68-72. [doi: [10.1177/002224298004400410](https://doi.org/10.1177/002224298004400410)]
39. Andrews FM, Withey SB. Developing measures of perceived life quality: Results from several national surveys. *Social Indicators Research.* 1974;1(1):1-26. [doi: [10.1007/bf00286419](https://doi.org/10.1007/bf00286419)]
40. Brand R, Ekkekakis P. Affective-Reflective theory of physical inactivity and exercise. *Ger J Exerc Sport Res.* 2017;48(1):48-58. [doi: [10.1007/s12662-017-0477-9](https://doi.org/10.1007/s12662-017-0477-9)]
41. Deci EL, Ryan RM. The "What" and "Why" of goal pursuits: human needs and the self-determination of behavior. *Psychological Inquiry.* 2000;11(4):227-268. [doi: [10.1207/s15327965pli1104\\_01](https://doi.org/10.1207/s15327965pli1104_01)]
42. Brock TC, Livingston SD. The Need for Entertainment Scale. In: *The Psychology of Entertainment Media: Blurring the Lines between Entertainment and Persuasion.* New Jersey. Lawrence Erlbaum Associates Publishers; 2004:255-274.
43. Deci EL, Ryan RM. Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology / Psychologie canadienne.* 2008;49(3):182-185. [doi: [10.1037/a0012801](https://doi.org/10.1037/a0012801)]
44. Gao W, Liu Z, Li J. How does social presence influence SNS addiction? A belongingness theory perspective. *Computers in Human Behavior.* 2017;77:347-355. [doi: [10.1016/j.chb.2017.09.002](https://doi.org/10.1016/j.chb.2017.09.002)]
45. Hatzinger LA, Do B, Hewus M, Crosley-Lyons R, Dunton GF. When physical activity fails to satisfy important psychological needs: Associations with engagement and mental health. *J Sports Sci.* 2025;43(24):3137-3145. [doi: [10.1080/02640414.2025.2574260](https://doi.org/10.1080/02640414.2025.2574260)] [Medline: [41090982](https://pubmed.ncbi.nlm.nih.gov/41090982/)]
46. Dunton GF, Do B, Crosley-Lyons R, Naya CH, Hewus M, Kanning M. Assessing basic and higher-level psychological needs satisfied through physical activity. *Front Psychol.* 2023;14:1023556. [FREE Full text] [doi: [10.3389/fpsyg.2023.1023556](https://doi.org/10.3389/fpsyg.2023.1023556)] [Medline: [36891201](https://pubmed.ncbi.nlm.nih.gov/36891201/)]
47. Bryant FB, Veroff J. *Savoring: A New Model of Positive Experience.* New York. Psychology Press; 2017.
48. Jose PE, Lim BT, Bryant FB. Does savoring increase happiness? A daily diary study. *The Journal of Positive Psychology.* 2012;7(3):176-187. [doi: [10.1080/17439760.2012.671345](https://doi.org/10.1080/17439760.2012.671345)]
49. Sato I, Jose PE, Conner TS. Savoring mediates the effect of nature on positive affect. *International Journal of Wellbeing.* 2018;8(1):18-33. [doi: [10.5502/ijw.v8i1.621](https://doi.org/10.5502/ijw.v8i1.621)]
50. Bryant FB, Chadwick ED, Kluwe K. Understanding the Processes that Regulate Positive Emotional Experience: Unsolved Problems and Future Directions for Theory and Research on Savoring. *International Journal of Wellbeing.* 2011;1(1). [FREE Full text] [doi: [10.5502/ijw.v1i1.18](https://doi.org/10.5502/ijw.v1i1.18)]
51. Radcliffe E, Lippincott B, Anderson R, Jones M. A pilot evaluation of mHealth app accessibility for three top-rated weight management apps by people with disabilities. *Int J Environ Res Public Health.* 2021;18(7):3669. [FREE Full text] [doi: [10.3390/ijerph18073669](https://doi.org/10.3390/ijerph18073669)] [Medline: [33915917](https://pubmed.ncbi.nlm.nih.gov/33915917/)]
52. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End user and implementer experiences of mHealth technologies for noncommunicable chronic disease management in young adults: systematic review. *J Med Internet Res.* 2017;19(12):e406. [FREE Full text] [doi: [10.2196/jmir.8888](https://doi.org/10.2196/jmir.8888)] [Medline: [29233804](https://pubmed.ncbi.nlm.nih.gov/29233804/)]
53. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth app usability questionnaire (MAUQ): development and validation study. *JMIR mHealth and uHealth.* 2019;7(4):e11500. [FREE Full text] [doi: [10.2196/11500](https://doi.org/10.2196/11500)] [Medline: [30973342](https://pubmed.ncbi.nlm.nih.gov/30973342/)]

54. House A, Bryant L, Russell A. Involvement in and experiences of the intervention and of research participation: results from a mixed-methods study. In: *Managing with Learning Disability and Diabetes: OK-Diabetes? a Case-Finding Study and Feasibility Randomised Controlled Trial*. Southampton, UK. NIHR Journals Library; 2018.
55. Oxman M, Chesire F, Mugisha M, Ssenyonga R, Nsangi A, Oxman AD, et al. et al. Participants' experiences of potential adverse effects of an intervention to improve critical thinking about health choices: a qualitative cross-trial process evaluation in Kenya, Rwanda and Uganda. *BMJ Open*. 2025;15(10):e104236. [FREE Full text] [doi: [10.1136/bmjopen-2025-104236](https://doi.org/10.1136/bmjopen-2025-104236)] [Medline: [41125290](https://pubmed.ncbi.nlm.nih.gov/41125290/)]
56. Alhubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. *J Multidiscip Healthc*. May 2016;9:211-217. [FREE Full text] [doi: [10.2147/JMDH.S104807](https://doi.org/10.2147/JMDH.S104807)] [Medline: [27217764](https://pubmed.ncbi.nlm.nih.gov/27217764/)]
57. Craft B, Carroll H, Lustyk M. Gender differences in exercise habits and quality of life reports: assessing the moderating effects of reasons for exercise. *Int J Lib Arts Soc Sci*. 2014;2(5):65-76. [FREE Full text] [Medline: [27668243](https://pubmed.ncbi.nlm.nih.gov/27668243/)]
58. Larsen S, Mozdoorzoy T, Kristiansen E, Nygaard Falch H, Aune TK, van den Tillaar R. A comparison of motives by gender and age categories for training at norwegian fitness centres. *Sports*. 2021;9(8):113. [doi: [10.3390/sports9080113](https://doi.org/10.3390/sports9080113)]
59. Gulec H, Smahel D. Individual and parental factors of adolescents' mHealth app use: nationally representative cross-sectional study. *JMIR mHealth and uHealth*. 2022;10(12):e40340. [FREE Full text] [doi: [10.2196/40340](https://doi.org/10.2196/40340)] [Medline: [36525286](https://pubmed.ncbi.nlm.nih.gov/36525286/)]
60. Klenk S, Reifegerste D, Renatus R. Gender differences in gratifications from fitness app use and implications for health interventions. *Mobile Media & Communication*. 2017;5(2):178-193. [doi: [10.1177/2050157917691557](https://doi.org/10.1177/2050157917691557)]
61. Lugones-Sanchez C, Recio-Rodriguez JI, Agudo-Conde C, Repiso-Gento I, G Adalia E, Ramirez-Manent JI, et al. EVIDENT 3 Investigators. Long-term effectiveness of a smartphone app combined with a smart band on weight loss, physical activity, and caloric intake in a population with overweight and obesity (Evident 3 Study): randomized controlled trial. *Journal of Medical Internet Research*. 2022;24(2):e30416. [FREE Full text] [doi: [10.2196/30416](https://doi.org/10.2196/30416)] [Medline: [35103609](https://pubmed.ncbi.nlm.nih.gov/35103609/)]
62. Damschroder LJ. Clarity out of chaos: use of theory in implementation research. *Psychiatry Research*. 2020;283:112461. [FREE Full text] [doi: [10.1016/j.psychres.2019.06.036](https://doi.org/10.1016/j.psychres.2019.06.036)] [Medline: [31257020](https://pubmed.ncbi.nlm.nih.gov/31257020/)]

## Abbreviations

**DTx:** digital therapeutics  
**EMA:** ecological momentary assessment  
**HRmax:** maximum heart rate  
**IRB:** Institutional Review Board  
**MCID:** minimal clinically important difference  
**MPA:** moderate intensity physical activity  
**ORBIT:** Obesity-Related Behavioral Intervention Trials  
**PA:** physical activity  
**REDCap:** Research Electronic Data Capture  
**RWE:** real-world evidence  
**SMART:** specific, measurable, actionable, realistic, timely (or time-bound)  
**SUS:** System Usability Scale

*Edited by A Kushniruk; submitted 19.Jan.2025; peer-reviewed by J Courtney, J-H Song; comments to author 11.Sep.2025; revised version received 01.Dec.2025; accepted 02.Jan.2026; published 20.Feb.2026*

### *Please cite as:*

Crosley-Lyons R, Hatzinger L, Hewus M, Wang W-L, Van Dyck D, Huh J, Hekler E, Dunton GF  
*An Early-Stage Digital Therapeutic Intervention to Enhance Affective Response During Physical Activity Among Adults With Overweight or Obesity: Benchmark-Driven Formative Testing Study*  
*JMIR Hum Factors* 2026;13:e71472  
 URL: <https://humanfactors.jmir.org/2026/1/e71472>  
 doi: [10.2196/71472](https://doi.org/10.2196/71472)  
 PMID:

©Rachel Crosley-Lyons, Lori Hatzinger, Micaela Hewus, Wei-Lin Wang, Delfien Van Dyck, Jimi Huh, Eric Hekler, Genevieve F Dunton. Originally published in *JMIR Human Factors* (<https://humanfactors.jmir.org>), 20.Feb.2026. 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 *JMIR Human Factors*, is properly cited. The complete bibliographic information, a link to the original publication on <https://humanfactors.jmir.org>, as well as this copyright and license information must be included.