The Impact of Mobile Technology-Delivered Interventions on Youth Well-being: Systematic Review and 3-Level Meta-analysis

Colleen S Conley1, PhD; Elizabeth B Raposa2, PhD; Kate Bartolotta2*, BA; Sarah E Broner1*, BA; Maya Hareli1, MA; Nicola Forbes2, MA; Kirsten M Christensen3, MA; Mark Assink4, PhD

1Department of Psychology, Loyola University Chicago, Chicago, IL, United States
2Department of Psychology, Fordham University, Bronx, NY, United States
3Department of Psychology, University of Massachusetts Boston, Boston, MA, United States
4Faculty of Social and Behavioural Sciences, University of Amsterdam, Amsterdam, Netherlands
*these authors contributed equally

Corresponding Author:
Colleen S Conley, PhD
Department of Psychology
Loyola University Chicago
1032 W. Sheridan Road
Chicago, IL, 60660
United States
Phone: 1 7735083603
Email: cconley@luc.edu

Abstract

Background: Rates of mental health problems among youth are high and rising, whereas treatment seeking in this population remains low. Technology-delivered interventions (TDIs) appear to be promising avenues for broadening the reach of evidence-based interventions for youth well-being. However, to date, meta-analytic reviews on youth samples have primarily been limited to computer and internet interventions, whereas meta-analytic evidence on mobile TDIs (mTDIs), largely comprising mobile apps for smartphones and tablets, have primarily focused on adult samples.

Objective: This study aimed to evaluate the effectiveness of mTDIs for a broad range of well-being outcomes in unselected, at-risk, and clinical samples of youth.

Methods: The systematic review used 5 major search strategies to identify 80 studies evaluating 83 wellness- and mental health-focused mTDIs for 19,748 youth (mean age 2.93-26.25 years). We conducted a 3-level meta-analysis on the full sample and a subsample of the 38 highest-quality studies.

Results: Analyses demonstrated significant benefits of mTDIs for youth both at posttest (g=0.27) and follow-up (range 1.21-43.14 weeks; g=0.26) for a variety of psychosocial outcomes, including general well-being and distress, symptoms of diverse psychological disorders, psychosocial strategies and skills, and health-related symptoms and behaviors. Effects were significantly moderated by the type of comparison group (strongest for no intervention, followed by inert placebo or information-only, and only marginal for clinical comparison) but only among the higher-quality studies. With respect to youth characteristics, neither gender nor pre-existing mental health risk level (not selected for risk, at-risk, or clinical) moderated effect sizes; however, effects increased with the age of youth in the higher-quality studies. In terms of intervention features, mTDIs in these research studies were effective regardless of whether they included various technological features (eg, tailoring, social elements, or gamification) or support features (eg, orientation, reminders, or coaching), although the use of mTDIs in a research context likely differs in important ways from their use when taken up through self-motivation, parent direction, peer suggestion, or clinician referral. Only mTDIs with a clear prescription for frequent use (ie, at least once per week) showed significant effects, although this effect was evident only in the higher-quality subsample. Moderation analyses did not detect statistically significant differences in effect sizes based on the prescribed duration of mTDI use (weeks or sessions), and reporting issues in primary studies limited the analysis of completed duration, thereby calling for improved methodology, assessment, and reporting to clarify true effects.

Conclusions: Overall, this study’s findings demonstrate that youth can experience broad and durable benefits of mTDIs, delivered in a variety of ways, and suggest directions for future research and development of mTDIs for youth, particularly in more naturalistic and ecologically valid settings.
Introduction

Youth Mental Health Needs

Rates of mental health problems among youth, including children, adolescents, and young adults, are alarmingly high and appear to have risen in recent decades [1,2]. Rates of impulse control disorders (eg, attention deficit hyperactivity disorder and conduct disorder) and some anxiety disorders begin rising as early as the age of 4 years, with sharp increases in the prevalence of anxiety, mood, and substance use disorders across adolescence and young adulthood [3]. Indeed, nationally representative samples of adolescents and young adults show that >40% of youth in this age range experience psychological disorders in a given year and lifetime prevalence rates are estimated to approach 60% [4,5]. Beyond diagnosable mental disorders, many youths struggle with a diverse array of subclinical emotional, behavioral, interpersonal, and academic challenges [6-10].

Despite the high prevalence of youth facing mental health problems, only one-third to one-half of those in need receive mental health treatment [11-14], and treatment rates are even lower among low-income youth [15] and those with marginalized racial and ethnic identities [13,16,17]. This treatment gap is also evident among college students [11], which is notable given that these youth often have convenient access to campus to no- or low-cost mental health services [18,19].

Youth and families face several barriers to receiving mental health services [20-24]. Many lack knowledge and awareness of common mental health problems and may assume that certain behavioral or emotional issues are simply temporary phases or difficulties they can address on their own. These types of assumptions may be compounded by the stigma about mental illness and psychological services within the youth’s family, culture, or broader community [25,26]. Many youths and caregivers also lack knowledge of available evidence-based treatments for mental health problems. Furthermore, several structural barriers severely limit the access of many youths and families to culturally sensitive, effective mental health care. Low-cost, evidence-based treatments for youth mental health problems are not available in many underserved communities across the world, including low-income rural and urban areas and countries with limited health care infrastructure. Even when such treatments are available, families may lack the time or resources needed to travel and take advantage of these treatments [27-30].

Technology’s Role in Youth Mental Health

Although efforts should continue to address the barriers that prevent formal services by qualified mental health professionals, it is also important to consider alternative ways of fulfilling the unmet mental health needs among youth. Mobile technology-delivered interventions (mTDIs), including mental health content delivered via mobile phones, tablets, and wearable smart devices (eg, watches, glasses, and virtual reality [VR] headsets), are potential ways of meeting this need. As of 2018, youth smartphone ownership and use were remarkably high, with 95% of teenagers having access to smartphones [31]. In 2016, the average age of first owning a smartphone was 10 years in the United States [32], and younger children commonly have access to smart devices through parents, siblings, or schools that provide tablets and other mobile devices to students. These mobile devices may be overlooked conduits for mental health information. Indeed, a recent survey of teenagers and young adults [33] reported that among those with moderate to severe depressive symptoms, 90% had searched the internet for information about mental health, and 38% used a mental health app. Parents also often use the internet for resources on health-related issues among their young children [34].

Technology can offer easy ways of connecting with mental health resources, such as mood-enhancing and skill-building apps purported to improve mental health. The ubiquitous, self-guided nature of such technology-delivered tools makes them appealing alternatives for those who are limited by access to, or trust in, formal mental health services [35]. Key themes in a recent review of research on internet-based help seeking for mental health difficulties among young people (aged up to 25 years) [36] showed that youth frequently engaged in technology-based (eg, internet-based) help seeking late at night (when traditional in-person mental health services are typically not available) and that youth endorsed several specific benefits of seeking help this way, including anonymity and privacy; lower perceived stigma and judgment; accessibility, including in times of crisis; and connection to others with similar experiences, which can foster a sense of community and acceptance.

Technology-Delivered Interventions for Youth Potential and Pitfalls

Although mTDIs have great potential to improve access to evidence-based mental health content, it is important to carefully evaluate their effects when it comes to mental health care, especially for youth. In contrast to computer- and internet-based technology-delivered interventions (TDIs), which are typically developed by clinicians and researchers to incorporate comprehensive and evidence-based treatment methods that parallel professional psychotherapy, mobile TDIs often lack such comprehensive, evidence-based principles while also introducing privacy and safety concerns [37]. The rapidly developing and competitive mobile app marketplace also poses some challenges in connecting evidence-based mental health practices with marketable and engaging mobile technology. Commercially available apps are typically designed by technology companies outside of the health care industry [38,39] with the aim of being engaging and attractive, thereby
prioritizing appealing features such as design and gamification over evidence-based clinical techniques [40,41]. In contrast, research-developed apps prioritize evidence-based content and rigorous trials, which slows widespread availability as it can take 2 decades from conceptualization to public dissemination [37,42], leaving a gap between research-tested and commercially available mTDIs. Indeed, recent reviews of the content of >10,000 purported mental health apps in the commercial marketplace have noted that a vast majority have not gone through rigorous intervention development and testing [43] and are lacking or inconsistent with evidence-based psychotherapy principles [44], including apps specifically targeting youth mental health [40,45]. Although research trials have demonstrated promising findings for some mTDIs, the heterogeneity and poor quality of certain mTDIs and studies have led to inconclusive evidence across outcomes [46], which makes the role of these interventions in mental health services less clear [47]. Finally, mental health technologies generally have low rates of engagement and adherence [35,43,48-50]. Thus, a key question for this emerging area of research is how mTDIs can best be designed, prescribed, and implemented to harness their benefits for youth.

Areas for Further Research

Previous reviews of TDIs, a broader category that goes beyond mobile interventions, have demonstrated the benefits of computer- and internet-based interventions, most commonly cognitive behavioral interventions, and most commonly examining the outcomes of depression, anxiety, and stress [51-58]. Similar to findings in adult populations, computer- and internet-based TDIs in youth samples have shown benefits, mostly in reducing internalizing symptoms, behavioral concerns, and eating disorders [51,52,54-60]. The literature on the efficacy of mobile TDIs, or mTDIs, is growing, with multiple meta-analytic reviews indicating positive impacts on a range of psychological outcomes in adults [42,53,61-63]. The emerging meta-analytic literature on mTDIs for youth is encouraging but limited, including generally beneficial results across (1) reviews blending a few trials of mTDIs together with mostly nonmobile TDIs in youth [52,64,65]; (2) a review combining 4 trials of mTDIs in children and adolescents with 21 trials of mTDIs in adults (mean age up to 59 years) [66]; (3) a recent meta-analysis of 12 youth trials, both with or without comparison groups, of smartphone apps exclusively on internalizing disorders [67]; and (4) another recent meta-analysis of 11 randomized controlled trials of smartphone apps for depression, anxiety, and stress in youth (aged 10-35 years) [68]. Although these initial findings are encouraging, a fitting next step for this emerging area of research is to meta-analytically review TDIs that are exclusively mobile in youth samples while including a broad array of youth clinical presentations and outcomes. Moreover, given the diverse designs of mTDIs for distinct youth characteristics and presenting problems, exploring the moderating influence of the interventions’ mobile technologies, theoretical orientations, technological and support features, and varying dosages would advance our ability to harness the full potential of mTDIs in improving youth well-being.

The Current Meta-analysis: Goals and Hypotheses

The current 3-level meta-analysis evaluated the impact of wellness- and mental health-focused mTDIs (including smartphones and tablets, other types of mobile phones, and other handheld and wearable devices, including mobile VR) for youth, broadly defined as children, adolescents, and intentional (eg, university student) young adult samples or those with a mean age of ≤26 years. Improving upon some limitations of previous reviews, we included published and unpublished reports, only included controlled (either randomized or quasi-experimental) designs, and evaluated a broad range of participant clinical presentations (eg, unselected, at-risk, or clinical samples), intervention theoretical orientations, and outcomes. Drawing on evidence from prior reviews, we predicted that these mTDIs would yield significant benefits at postintervention on diverse indicators of youth well-being relative to comparison conditions. In addition, we examined the role of several potential moderators of intervention impact within the categories of methodological, youth, and intervention characteristics.

Methodological Characteristics

Timing of Outcomes

Prior reviews note that there are a limited number of studies assessing the long-term effects of TDIs [51,54] and mTDIs [68] on youth. The reviews that compare the effects at postintervention versus later follow-up periods have been mixed, with some finding that effects are stable into follow-up periods (eg, parenting TDIs [59]) and others finding that some or all effects diminish over time (eg, adult mTDIs [69] and parenting TDIs [70]). Given that youth might have added challenges in implementing long-term gains [57], we tentatively predicted that the timing of the outcome assessment (posttest vs follow-up) would moderate the strength of the mTDIs’ effects such that the effects of mTDIs would wane over time.

Outcome Type

Prior reviews have established the benefits of mTDIs in reducing depression, anxiety, and stress, mostly in adults [42,62,69], with emerging evidence in youth [67,68]. In addition, mTDIs have been shown to be effective in improving life satisfaction, quality of life, and psychological well-being [69]. To broadly evaluate the potential impact of mTDIs on youth, we examined a broad range of youth outcomes, including those that have not yet been examined in prior reviews. Therefore, we expected mTDIs to have a beneficial impact on depression, anxiety, stress, and well-being, and explored whether mTDIs would also have beneficial effects on other outcome types, such as psychosocial strategies and skills, interpersonal relationship factors, academic functioning, health-related behaviors, or knowledge.

Comparison Group Type

Reviews of TDIs and mTDIs have generally found that effects are largest when they are compared with no-intervention or wait-list groups and smaller when compared with groups that are more active and clinically potent [42,52,58,60,62,69-71]. Thus, we predicted that the comparison group type would moderate the effects of mTDIs. Specifically, we expected that mTDIs would demonstrate the strongest benefits compared with
no intervention (eg, wait-list), followed by inert interventions (including information-only and attentional or placebo controls), and demonstrate noninferiority compared with clinical comparisons, including usual clinical care and established clinical interventions.

Youth Characteristics

Age
Several previous reviews of TDIs in youth have demonstrated that older participants experience a greater reduction in symptoms than their younger counterparts [51,55,65,72]; however, others have found no effect of age [57,59], and preliminary evidence on a small sample of mTDIs in adolescents and young adults also failed to find an effect of age [68]. In an exploratory fashion, we examined whether age moderates the effects of mTDIs.

Gender
The few reviews on the mental health benefits of smartphone apps that explored the role of participants’ gender have revealed nonsignificant effects in adult [42] and youth [59] samples. Nevertheless, given the differences in the rates of various mental health problems as a function of gender across development [73,74], we tested the effects of gender in an exploratory fashion.

Risk Level
Some prior reviews of mTDIs with adult samples have indicated that higher pretreatment severity (ie, clinical diagnosis or elevated mental health symptoms) is related to a greater reduction in symptoms and, therefore, produces a larger effect size (ES) [60,63]. In contrast, Pennant et al [54] found greater effects of computerized therapies for youth with subclinical symptoms versus a clinical diagnosis of anxiety; however, this effect was not found for depression. Two more recent reviews suggest inconclusive evidence regarding whether TDIs or mTDIs are more effective for youth who present with diagnoses or severe symptomatology [47,65]. Therefore, we examined participant risk level (ie, clinical diagnosis, elevated symptoms, and nonclinical sample) as an exploratory moderator.

Intervention Characteristics

Type of Technology
A prior review on the impact of mTDIs on both youth and adults did not find significant differences in effects by type of technology (ie, smartphone apps/products versus other types of mobile phones, PDAs, wearable devices, or VR headsets) but noted that smartphone apps produced (nonsignificantly) larger ESs than both PDA and SMS text messaging interventions [66]. Given the limited information, we tested the impact of technology type as an exploratory moderator.

Guiding Theoretical Framework
Cognitive behavioral–based and mindfulness- or acceptance-based theoretical frameworks are the most commonly examined theoretical frameworks among TDIs for youth and adults [51,54-58] and mTDIs for adults [42,53,62,69] and have generally yielded positive effects for problems such as anxiety, depression, externalizing behaviors, and quality of life. Thus, we hypothesized that mTDIs with cognitive behavioral–based and mindfulness- or acceptance-based theoretical frameworks would produce significant effects and examined the impact of additional theoretical orientations (eg, motivational or positive psychology) in a more exploratory fashion.

Technological and Support Features
Prior reviews have suggested that specific features of mTDIs, such as tailoring (ie, content shifting based on responses), gamification, and automatic reminders, may increase engagement and yield more robust effects [53,75]; however, much of this research has been conducted with adult samples. Therefore, we explored the potential moderating impact of various technological features such as personalization, tailoring, gamification, and social elements (eg, peer mentoring).

Research also points to the possibility that support from a human or virtual (“bot”) professional, who can provide guidance, coaching, accountability, and, in some cases, supervised skills practice, may lead to increased adherence to mTDIs [47,76,77]. Indeed, several reviews have indicated that self-guided interventions are generally more effective with some access to human support or guidance, in part because of increased engagement and adherence [53,78,79]. Similarly, research on psychotherapy and other in-person youth interventions have highlighted the benefits of supervised practice in contributing to youths’ psychological skill development, especially when delivered over multiple sessions [79-84]. Nevertheless, the overall evidence on human support for mTDI use is mixed, with a handful of reviews not finding added benefits for interventions incorporating human support as compared with those that do not [42,54,55,85]. Thus, we explored the potential impact of human or bot support elements, such as coaching, supportive accountability, and supervised skills practice.

Dosage: Frequency and Duration
Previous studies have rarely investigated the dose-response relationship between TDI use and outcomes, and those that have done so tend to yield mixed results [47,51,54]. Some reviews have found that higher dosages, or longer durations, predicted a greater reduction in symptoms with particular outcomes (eg, problem behavior or depression) [59,86]. However, other studies were unable to establish such a relationship [57,85]. Therefore, we explored whether the prescribed or completed frequency (eg, weekly) or duration (eg, minutes, sessions, or weeks) of the intervention moderated the benefits of mTDIs.

Methods

Search Strategy and Study Selection
We used 5 systematic search strategies to assemble an unbiased, representative sample of published and unpublished controlled trials. First, we conducted searches for reports appearing through March 2021 in 5 academic databases: PsycINFO, ERIC, ProQuest Digital Dissertations, MEDLINE (Web of Science), and PubMed. We used a combination of several groups of search terms to find studies meeting our criteria for (1) participants (eg, child*, adolescence*, teen, youth, young adult, university student), (2) interventions (eg, mental health, psychological intervention, cognitive behavioral, mindfulness), (3) mental health (eg, depress*, anxiety*, well-being), (4) technology (eg,
smartphone app*, mobile app*, tablet-based), and (5) research design (eg, RCT, controlled trial, clinical trial, quasi, comparison group, PRISMA). Second, we also inspected the reference lists of each study meeting our criteria and of relevant previous reviews. Third, we hand-searched the contents of 16 selected journals most likely to publish studies on mobile mental health interventions involving youth. Fourth, we hand-searched the contents of proceedings for the recent years of 7 relevant academic conferences. Finally, we contacted authors of prior reviews, reports, and conference proceedings relevant to our sample to inquire about additional published or unpublished evaluations of fitting trials. Further details on these search strategies are provided in Multimedia Appendix 1.

To be included in our final sample, the studies had to meet six criteria: (1) examine an automated psychological or behavioral intervention, either selecting participants based on a diagnosis or risk factor or targeting an unselected sample to promote mental health and wellness; (2) deliver the intervention primarily via mobile (handheld or wearable) technology, including pre–cellular technology handheld computers (eg, palm pilot and PDA), mobile cellular phones or tablets (eg, iPad and iPod touch)—using SMS text messaging, instant messaging, or more current mobile mental health apps—and wearable devices (eg, smart watch, smart glasses, VR headsets that are fully self-contained or linked with a mental health app on a smartphone or tablet device that is portable and able to be used in the participant’s home); (3) contain at least one quantitatively assessed mental or behavioral health outcome measure (described in the following sections) for which ESs could be calculated; (4) target youth, broadly defined as children, adolescents, and intentional (eg, university student) young adult samples or those with a mean age ≤26 years (including interventions delivered to parents that targeted youth outcomes); (5) include a comparison group with at least 10 participants assigned to each condition; and (6) be reported in English, Spanish, Dutch, or German.

We excluded interventions with a primary focus on academics or physical health (eg, nutrition, weight loss, or diabetes management) but included studies that focused on psychobehavioral health such as smoking cessation, insomnia, and disordered eating. We did not include interventions delivered through audio or video tapes or videodiscs, a local computer program, or a website only. In addition, we did not include mobile interventions that were primarily reliant on human support (eg, therapists sending messages). Finally, we excluded interventions comprising solely medication reminders.

Figure 1 shows a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of sample searching, selection, and inclusion. The aforementioned search procedures identified 7487 potentially relevant reports, including 2353 (31.43%) duplicates that were removed. An additional 51.44% (3851/7487) of reports were eliminated as they did not meet our inclusion criteria. Among the 1283 eligible reports, some contained variants of the same intervention (eg, 2 interventions with the same active component but varying lengths), and we only included the intervention that was more comprehensive (ie, contained more elements or was longer in duration) or completely technology-based. However, if conceptually distinct interventions (eg, 2 different apps using different techniques) were evaluated in the same report, each intervention was coded separately. Data from multiple reports on the same sample and intervention were combined into a single report, reducing 16 overlapping reports to a sample size of 6.

In cases where means and SDs were not included in the original reports or effects could not be calculated because of insufficient data, we attempted to contact study authors to secure missing data. On the basis of a lack of author response, we excluded 16 studies for which no ESs could be calculated for any relevant outcome measure. This screening process led to a final sample of 83 interventions reported in 80 studies between 2005 and 2021.
Study Coding (Data Extraction)

**Methodological Characteristics**

For each report, we coded the year of the report, publication status, country in which the intervention took place, type of experimental design and comparison group, sample size, outcome types, and additional codes described in the following sections.

**Timing of Outcomes**

We coded the number of weeks between pre- and postintervention outcome assessments and between postintervention and each follow-up assessment period.

**Outcome Type**

We coded a broad array of outcomes to capture the various psychosocial and related aspects of functioning that might be affected by mTDIs. The relevant outcomes assessed in our sample of studies were classified into 14 possible categories, some of which were conceptually nested under higher-order categories, as noted in Textbox 1.
Textbox 1. Outcome types coded.

- General psychological well-being or distress included 2 subcategories:
  - **Stress** (eg, perceived and physiological indices of stress)
  - **General or global psychological distress** and well-being (eg, distress, positive and negative affect, mood states, quality of life, happiness, or life satisfaction)

- Psychosocial strategies or skills included 2 subcategories:
  - **Social-cognitive strategies or styles** (eg, different types of affective, cognitive, and social skills related to effective coping strategies, help-seeking behaviors, or mindfulness practices; overcoming dysfunctional beliefs, rumination, or hostility; resilience; or emotional self-awareness and regulation)
  - **Self-perceptions** (eg, self-esteem or self-efficacy)

- Internalizing symptoms included 2 subcategories:
  - **Depression**
  - **Anxiety**

- **Other (noninternalizing) mental health problems** (eg, autism, attention deficit hyperactivity disorder, or eating disorders)

- **Health and health behavior** (eg, substance use, sleep, physical activity, pain, or eating behaviors)

- **Interpersonal relationships** (eg, conflict, perceived social support, belongingness, loneliness, or social skills)

- **Academics** (eg, academic performance or adjustment)

- **Psychology or health-related knowledge** (eg, knowledge about topics such as substance use norms and consequences or sleep hygiene)

- **Psychosocial outcomes in someone other than the target youth** (eg, parent stress, warmth, use of punishment)

- **Other** (eg, perceptions of productivity, stigma, or close friend’s smoking behavior)

- **Intervention (ie, app) ratings** (eg, intervention feasibility and social validity, acceptability of the mobile technology–delivered intervention (mTDI), and its uptake or use)

### Comparison Group Type

Studies were coded as having 1 of 3 different comparison groups. The majority of studies included a no-intervention (eg, wait-list control) condition in which the comparison group only completed assessment procedures. Some studies compared the intervention of interest with an inert comparison group, whether information-only (eg, pamphlets or website links to general health-related information), attention-placebo (eg, passive SMS text messages), or minimal treatment-as-usual (eg, standard protocol before a medical procedure) conditions that did not contain the therapeutic elements of the evaluated intervention. These comparison groups generally attempted to control for nonspecific factors such as attention or social interaction. Finally, some studies included a clinical comparison group, whether a usual clinical care comparison or some other established (validated or otherwise intended to be beneficial) intervention. In some studies with a clinical comparison group, both the mTDI and comparison groups received a similar base intervention (eg, counseling vs counseling+app) and thus tested the added or incremental benefit of the mTDI of interest.

### Youth Characteristics

**Age, Gender, Race, and Ethnicity**

When the information was available, studies were coded for the sample age (mean, SD, and range), gender, race, and ethnicity.

**Risk Level and Type**

We coded whether researchers selected participants based on particular symptoms or risk factors into the following categories: (1) psychological clinical sample (ie, symptoms indicative of a Diagnostic and Statistical Manual of Mental Disorders diagnosis) [87,88], (2) psychological or mental health at-risk sample (ie, subclinical symptoms of psychopathology), (3) nonmental health risk (ie, medical risk, diagnosis, or procedure), and (4) general (unselected) community sample not selected for any particular risk factor.

### Intervention Characteristics

**Type of Technology**

We coded each intervention’s primary and secondary (if relevant) type of technology into one of the following categories: (1) smartphone or tablet (eg, iPhone, iPad, or iPod touch), (2) presmartphone mobile device (eg, presmart mobile phone, palm pilot, or PDA), (3) mobile VR (eg, headsets) or video game (ie, handheld), and (4) other wearable devices (eg, smart watch, biosensor or activity monitor, and smart glasses). VR headsets and other wearable devices were typically used in conjunction with smartphones or tablets. Finally, some interventions were also able to be accessed on a computer as a secondary type of technology.

### Guiding Theoretical Framework

Interventions were coded as having one of the following primary guiding theoretical frameworks: (1) cognitive behavioral, (2) [89].
mindfulness- or acceptance-based (eg, mindfulness-based stress reduction, mindfulness-based cognitive therapy, acceptance and commitment therapy, or dialectical behavior therapy), (3) blended cognitive behavioral and mindfulness, (4) other or multiple (ie, positive psychology, interpersonal, motivational or stages of change, and transtheoretical), or (5) atheoretical or not specified. When mTDIs were available in the commercial market, they were consulted directly to supplement the information obtained from the research reports.

Technological and Support Features

We coded whether the intervention included personalization (ie, the ability to alter the app environment through features such as personal preferences; personal dashboards; or use of photos, music, or contacts), tailoring (ie, the use of algorithms that alter intervention content based on contact sensing, prior responses, feedback, or other input), a social component (eg, forum or social media use or mentoring), or gamification (eg, rewards, badges, points, levels, or quests).

We also coded several intervention features designed to support participants in using the mTDI: (1) training or orientation for the participants about using the mTDI (eg, virtual training within the app or via email or video chat, in-person training, or a paper manual); (2) in-person element besides orientation or training (eg, simultaneous counseling); (3) reminders sent to encourage the use of the mTDI, either automatically through the app (eg, push or banner notifications) or outside of the app (eg, emails, texts, or calendar reminders); (4) human or bot (automatic) support (eg, supportive SMS text messages, phone calls, or personalized feedback) specifically around the mTDI; (5) targeted guidance indicative of supportive accountability, designed to increase adherence to an intervention via support and accountability from a trustworthy coach who assists with setting process-oriented expectations and goals [77]; and (6) targeted guidance in the form of supervised skills practice [79].

Dosage: Frequency and Duration

When the information was available, studies were coded for the intervention’s prescribed and completed (both objectively determined and self-reported) frequency and duration. Specifically, the frequency of use was coded as one of the following categories: at least 4 days per week or as much as feasible, 2 to 3 days per week, once per week, less than once per week, one-time session, and not stated or at user discretion. The intervention duration was coded in terms of minutes, weeks, and sessions. When data were available, we also calculated the percentage of the completed duration of the intervention by dividing the completed duration by the prescribed duration.

Risk of Bias (Quality) Assessment

For study quality, we followed the approach of an integrative study quality coding scheme [89] designed to draw upon the strengths of several previously validated quality indices, including the Cochrane Collaboration’s tool for assessing the risk of bias [90-92]. This coding scheme rates each study on 10 features: peer review and impact factor, experimental design, sample size, attrition, reliability of measures, validity of measures, adjustment for pretest differences, intent-to-treat analysis, reporting of sample characteristics, and involvement of study authors in mTDI development. Each feature is rated on a 4-point scale (from 0, indicating the lowest quality, to 3, indicating the highest quality). The 10 item scores are then summed, resulting in a score for which a score of 20 represents average or normal research practices.

Reliability of Coding

A team of 5 trained postbaccalaureate and graduate students assessed the studies for eligibility and inclusion in the meta-analysis and met weekly to review any questions for consensus. A team of 6 graduate students with advanced clinical and quantitative training then reviewed and coded eligible reports for descriptive features, moderators, quality indicators, and outcome data. The coders were supervised by 3 faculty members with expertise in clinical psychology, mTDIs, and meta-analytic procedures. After the iterative training phase, coders had ongoing opportunities for consensus checks through a consultation system and weekly faculty supervision. From a subsample of 18 to 31 studies (depending on the code) containing 44 interventions, 46 comparisons, and 108 ESs, any code that did not reach adequate reliability (ie, >0.80 k, 85% agreement, or 0.95 intraclass correlation coefficient, as fitting) [93,94] was reviewed by at least one other coder in the entire sample. Lead authors provided an additional review of randomly selected articles throughout the coding process. Any questions or discrepancies were resolved through discussions.

Meta-analytic Strategy

ES Calculation

Cohen $d$ was calculated for each outcome to reflect the effect of mTDIs relative to the comparison condition, with positive ESs representing outcomes in which the intervention group outperformed the comparison group. If $d$ values could not be obtained directly from primary studies, the formulas by Borenstein et al [95] and Lipsey and Wilson [96] were used to transform the reported statistical information into Cohen $d$. Whenever possible, $d$ values were calculated using means and SDs, frequencies or proportions, odds ratios, or results from $F$ or $t$ tests. If a primary study did not report sufficient information to extract or calculate the ES, the study authors were contacted for additional information. When the only information available indicated that an ES was nonsignificant, we conservatively set that ES to zero, following Mullen [97]. This procedure was preferred above excluding primary studies from the review, as the latter would reduce the statistical power in the analyses. To correct for pretreatment differences, we adjusted the postintervention and follow-up effects for presubmission baseline outcome levels (using subtraction, similar to procedures in other meta-analyses) [89,98,99] when pretreatment data were available. Finally, before analysis, all ESs were converted to Hedges $g$ to account for potential bias in small sample sizes.

The 3-Level Meta-analytic Model

Most primary studies included in this review reported on multiple intervention effects, typically because multiple outcomes were tested or multiple comparison conditions were part of the study design. The resulting dependency in ESs (ie, the fact that ESs extracted from the same study are more alike than the ESs extracted from different studies) violates the
assumption of independent ESs underlying traditional meta-analytic techniques [96]. Therefore, a 3-level random-effects model was used for all analyses [100-104]. In this 3-level model, 3 sources of variance were modeled: sampling variance of the observed ESs (ie, sampling variance; level 1), variance between ESs derived from the same study (ie, within-study variance; level 2), and variance in ESs derived from different studies (ie, between-study variance; level 3). The sampling variance at level 1 of the model is not estimated but considered known and calculated using the formula given by Cheung [101].

To determine whether testing select moderators would be informative, we first examined the ES heterogeneity by testing the significance of the within-study variance (level 2) and the between-study variance (level 3). We performed 2 one-sided log-likelihood ratio tests in which the deviance of the full model was compared with the deviance of the model without one of these variance parameters. If the within-study variance or the between-study variance were significant, we proceeded with the moderator analyses. The coded variables were only tested as moderators when (categories of) these variables were based on at least three studies or three ESs. In some cases, we consolidated categories with <3 studies or ESs into another (or the other) category.

Software and Parameters

We used the function rma.mv of the metafor package [105] in the R statistical environment (version 3.6.1; R Foundation for Statistical Computing) [106], following the setup and R syntax by Assink and Wibbelink [100], to model the 3 sources of ES variance [103,104]. The overall effect was estimated using an intercept-only model, and potential moderators were examined by adding these variables as covariates to the intercept-only model. The t distribution was used in testing individual regression coefficients of the models and for calculating the corresponding Cs [107]. When models were extended with categorical moderators comprising >3 categories, the omnibus test followed an F distribution. The restricted maximum likelihood estimation method was used to estimate the model parameters. Before conducting moderator analyses, continuous variables were centered on their means, and dichotomous dummy variables were created for categorical variables. The log-likelihood ratio tests were conducted as 1-tailed, whereas all other significance tests were conducted as 2-tailed. The significance level was set to 0.05 in all analyses, and 95% Cs were estimated.

Publication Bias

A problem that may arise in meta-analysis is the file drawer problem [108], in that studies with nonsignificant or negative results are less likely to be published than studies that produced significant and positive results. To reduce this problem, we attempted to be exhaustive in our search strategy to retrieve both published and unpublished primary studies. To further assess bias in our data set of ESs, 2 analyses were conducted. First, we performed the trim-and-fill analysis by Duval and Tweedie [109,110] to examine the symmetry of a funnel plot in which ESs were plotted against their SEs. In the case of publication bias, the plot is asymmetrical, as the ESs are missing to the left of the estimated mean. The trim-and-fill algorithm estimates these missing ESs using an iterative nonparametric method. After imputing these ESs, the symmetry of the plot is restored, and an adjusted overall effect can be estimated. We also examined bias by performing the Egger test, in which ESs are regressed on their SEs [111]. This was performed by adding the SE as a covariate to an intercept-only, 3-level meta-analytic model. In this model, a significant positive slope indicated the presence of publication bias.

Results

Study Sample and Descriptive Characteristics

Multimedia Appendix 2 [112-200] provides a table with details about each of the 80 studies eligible for this meta-analysis, 3 (4%) of which contained 2 eligible interventions and 10 (13%) that contained 2 eligible comparison groups, yielding 83 interventions, 93 comparisons, and a combined sample size of 19,748 youth. Of these 80 studies, 76 (95%) provided estimates of 484 postintervention ESs, and 29 (36%) studies provided estimates of 225 follow-up ESs.

Several aspects of the 80 included studies are worthy of comment. First, most (68/80, 85%) of the studies were published in peer-reviewed scientific journals, and the remainder were unpublished dissertations or in-preparation manuscripts that extended prior peer-reviewed work published as a pilot trial or presented at an academic conference. In addition, most of the studies were published within the past decade or so, with 96% (77/80) published since 2010 and 28% (22/80) since 2020. Of the 80 studies, 33 (41%) were conducted in the United States, with 36 (45%) reports from the broader North American continent, 23 (29%) from Europe, 11 (14%) from Australia, 9 (11%) from Asia or the Middle East, and 1 (1%) from South America.

In terms of participants, across the 93% (74/80) of studies reporting relevant demographic information (and among the 67/80, 84% of studies reporting SD), the average age ranged from 2.93 to 26.25 (weighted mean 15.92; SD 2.86) years, and on average, 63.83% of study samples were female (but notably, most studies did not report on, or likely assess, gender other than female or male). Only 38% (30/80) of studies provided a full breakdown of participant race and ethnicity, and 23% (18/80) provided no information on these demographics at all. Furthermore, 63% (50/80) of studies selected participants based on one or more risk factors versus recruiting a general community sample. The most common risk factor used for participant recruitment and screening was subclinical psychological risk (eg, substance use or elevated depression; 30/80, 38% of studies), followed by some nonmental health risk (12/80, 15% of studies; in all cases within this sample, this was a medical diagnosis such as spina bifida or obesity or a medical procedure such as surgery or dental work), and, finally, participants with a clinical psychological or psychiatric diagnosis (eg, anxiety or autism; 8/80, 10% of studies).

In terms of the 83 interventions, 74 (89%) used smartphones or tablets (1 used an iPod touch); 4 (5%) used presmartphone
mobile devices (all phones, including Motorola A925, Sony Ericsson, and Vodafone); 4 (5%) used VR headsets, either freestanding or in conjunction with a mobile phone app; and 1 (1%) used a handheld video game. Most (70/83, 84%) of the interventions took place in participants’ daily environments; however, several (13/83, 16%) took place in a medical setting (eg, to address anxiety or pain related to a medical procedure). The most prevalent guiding theoretical framework of the mTDIs was cognitive behavioral (36/83, 43% of the interventions), followed by other or transtheoretical frameworks (eg, positive psychology and motivational; 20/83, 24%), mindfulness- or acceptance-based (17/83, 20%), and a few atheoretical or unspecified frameworks (3/83, 4%). Furthermore, in nonexclusive categories, the interventions’ technological features included personalization (18/83, 22%), tailoring (36/83, 43%), a social component (10/83, 12%), and gamification (20/83, 24%). In terms of support features, of the 83 interventions, 30 (36%) included some sort of orientation or training (either virtual or in person), 12 (14%) contained information about how much participants engaged in actually or completed dosage (or both) of their interventions, whereas the remainder of the comparisons involved either passive information-only or placebo groups (28/80, 35%) or, less commonly, clinical comparisons that were intended to have therapeutic benefits (15/80, 19%). For the studies (43/80, 54%) that used an active (inert or clinical) comparison, the modality of the comparison group was distributed fairly evenly across in-person (20/43, 47%) and other technology-based interventions (18/43, 42%), with just a few (5/43, 12%) having some other modality (ie, blended interventions containing both technology and in-person elements or paper-and-pencil materials).

All (80/80, 100%) studies provided some information about the prescribed or completed dosage (or both) of their interventions, whether objectively pulled from the mTDI or self-reported by the participants; however, the specific details reported were variable. Of the 83 interventions, 13 (16%) were single-session interventions and the remainder were prescribed to range from 4 to 2505 sessions (weighted mean 89.86, SD 374.25; k=42 studies reporting on 44 interventions) across a time span of 2 days to 43.45 weeks (weighted mean 7.48, SD 7.46; k=65 studies reporting on 68 interventions). Of these 70 interventions (contained in 67 studies), 43 (61%) were prescribed for daily use, 9 (13%) for 2 to 3 days per week, 7 (10%) for once a week, and the remaining 11 (16%) were either prescribed to be used as needed or at the user’s discretion or not stated in the report. In terms of duration of use, the prescribed minutes of use for interventions ranged from 5 to 3650 minutes (mean 345.25, SD 789.95; k=32 studies reporting on 32 interventions). Notably, only 47 out of 80 (59%) studies provided some sort of objective information about how much participants actually engaged in the intervention (eg, number of sessions, minutes, or weeks). Using all available information, we calculated the intervention completion percentage and found the average to be 85.05% of the researchers’ prescribed sessions (k=29 studies reporting on 30 interventions), 87.19% of the prescribed intervention minutes (k=13 studies reporting on 13 interventions), and 86.7% of the prescribed intervention weeks (k=37 studies reporting on 43 interventions).

Studies assessed a variety of psychosocial outcomes, which we originally coded in 14 categories (see the Methods section) and then consolidated into 6 categories because of conceptually similar content or small numbers of studies or effects (see the final list of consolidated categories in the note below Multimedia Appendix 2).

Notably, 25% (20/80) of studies also provided information about the intervention group’s ratings on measures of the intervention’s social validity (eg, user satisfaction, perceived usefulness, quality, usability, or acceptability). As these data were generally only available at the posttest time points and for intervention but not comparison groups, we do not report ESs on these types of outcomes. However, to analyze trends in diverse measures of mTDI social validity across all studies with such data available, we standardized all available Likert scale ratings for these constructs onto a single scale, with 0 representing the lowest and 100 the highest possible rating of social validity. On this standardized scale, the average rating (weighted by included sample size) for self-report scores of the mTDIs’ social validity was 58.24 and ranged quite widely across studies (30.20-100).

There was variability in the types of comparison groups as well. Slightly less than half (37/80, 46%) of the comparisons involved groups such as wait-lists that contained no active intervention, whereas the remainder of the comparisons involved either passive information-only or placebo groups (28/80, 35%) or, less commonly, clinical comparisons that were intended to have therapeutic benefits (15/80, 19%). For the studies (43/80, 54%) that used an active (inert or clinical) comparison, the modality of the comparison group was distributed fairly evenly across in-person (20/43, 47%) and other technology-based interventions (18/43, 42%), with just a few (5/43, 12%) having some other modality (ie, blended interventions containing both technology and in-person elements or paper-and-pencil materials).

### Average Effect of mTDIs

The average ES across all possible comparisons within the 80 studies (yielding 709 ESs across posttest and follow-up assessments) was g=0.27 (P<.001; 95% CI 0.20-0.33). There was significant heterogeneity across studies (σ² level 3=0.06; P<.001; 51.76% of the variance among ESs), as well as between ESs extracted from the same study (σ² level 2=0.03; P<.001; 27.50% of the variance among ESs). Random sampling error accounted for 20.74% of the variance. To explore the substantial variability between and within studies, a number of moderators were considered. These analyses are described in the following 3 sections and detailed in Table 1.
Table 1. Moderators of the effectiveness of mobile technology–delivered interventions for youth.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>$k^b$</th>
<th>Effect sizes, n</th>
<th>$B_0$ (intercept), g (95% CI)</th>
<th>$B_1$ (slope), g (95% CI)</th>
<th>$F$ (df$_1$, df$_2$)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Methodological characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study quality</td>
<td>80</td>
<td>709</td>
<td>0.25 (0.19 to 0.31)$^{***}$</td>
<td>-0.02 (–0.04 to –0.01)$^{**}$</td>
<td>7.03 (1, 707)</td>
<td>.01$^c$</td>
</tr>
<tr>
<td><strong>Timing of outcome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest (RC$^{cd}$)</td>
<td>76</td>
<td>484</td>
<td>0.27 (0.21 to 0.33)$^{***}$</td>
<td>N/A$^a$</td>
<td></td>
<td>.07</td>
</tr>
<tr>
<td>Follow-up</td>
<td>29</td>
<td>225</td>
<td>0.26 (0.19 to 0.34)$^{***}$</td>
<td>-0.01 (–0.06 to 0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outcome type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General psychological well-being or distress (RC)</td>
<td>35</td>
<td>98</td>
<td>0.28 (0.20 to 0.37)$^{***}$</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internalizing (depression, anxiety)</td>
<td>44</td>
<td>145</td>
<td>0.30 (0.22 to 0.39)$^{***}$</td>
<td>0.02 (–0.06 to 0.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (noninternalizing) mental health</td>
<td>7</td>
<td>42</td>
<td>0.21 (0.04 to 0.38)$^*$</td>
<td>-0.07 (–0.25 to 0.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychosocial strategies and skills</td>
<td>26</td>
<td>161</td>
<td>0.34 (0.25 to 0.42)$^{***}$</td>
<td>0.05 (–0.02 to 0.13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health (behavior; eg, substance use)</td>
<td>35</td>
<td>190</td>
<td>0.24 (0.16 to 0.32)$^{***}$</td>
<td>-0.04 (–0.15 to 0.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (eg, knowledge or relationships)</td>
<td>20</td>
<td>73</td>
<td>0.15 (0.04 to 0.25)$^{**}$</td>
<td>-0.14 (–0.25 to –0.03)$^{**}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Comparison group type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No intervention (RC; eg, wait-list)</td>
<td>41</td>
<td>376</td>
<td>0.28 (0.20 to 0.36)$^{***}$</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inert (eg, placebo or information-only)</td>
<td>30</td>
<td>186</td>
<td>0.26 (0.18 to 0.35)$^{***}$</td>
<td>-0.02 (–0.11 to 0.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical (eg, established intervention)</td>
<td>18</td>
<td>147</td>
<td>0.24 (0.12 to 0.36)$^{***}$</td>
<td>-0.04 (–0.17 to 0.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean age (years)</td>
<td>74</td>
<td>669</td>
<td>0.26 (0.19 to 0.32)$^{***}$</td>
<td>0.003 (–0.01 to 0.01)</td>
<td>0.17 (1, 667)</td>
<td>.68$^c$</td>
</tr>
<tr>
<td>Gender (percentage female)</td>
<td>77</td>
<td>699</td>
<td>0.26 (0.20 to 0.33)$^{***}$</td>
<td>0.002 (–0.002 to 0.01)</td>
<td>1.06 (1, 697)</td>
<td>.30</td>
</tr>
<tr>
<td><strong>Risk level and type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General sample not selected for risk (RC)</td>
<td>30</td>
<td>247</td>
<td>0.19 (0.09 to 0.29)$^{***}$</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonmental health (ie, medical) risks</td>
<td>12</td>
<td>46</td>
<td>0.52 (0.33 to 0.72)$^{***}$</td>
<td>0.33 (0.11 to 0.55)$^{**}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological or mental health at-risk sample</td>
<td>30</td>
<td>332</td>
<td>0.29 (0.19 to 0.39)$^{***}$</td>
<td>0.09 (–0.05 to 0.23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological clinical sample (diagnosis)</td>
<td>8</td>
<td>84</td>
<td>0.20 (0.01 to 0.39)$^*$</td>
<td>0.01 (–0.21 to 0.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intervention characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary type of technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.82 (2, 706)</td>
<td>.44</td>
</tr>
<tr>
<td>Smartphone or tablet (RC)</td>
<td>71</td>
<td>665</td>
<td>0.25 (0.19 to 0.32)$^{***}$</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presmartphone mobile device</td>
<td>4</td>
<td>23</td>
<td>0.37 (0.09 to 0.65)$^*$</td>
<td>0.11 (–0.18 to 0.40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile VR$^{f}$ headset or handheld video game</td>
<td>5</td>
<td>21</td>
<td>0.41 (0.13 to 0.68)$^{**}$</td>
<td>0.15 (–0.13 to 0.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Guiding theoretical framework</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.04 (4, 704)</td>
<td>.38</td>
</tr>
<tr>
<td>Cognitive or behavioral (RC)</td>
<td>35</td>
<td>249</td>
<td>0.31 (0.21 to 0.40)$^{***}$</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mindfulness or acceptance</td>
<td>16</td>
<td>238</td>
<td>0.28 (0.15 to 0.42)$^{***}$</td>
<td>-0.02 (–0.19 to 0.15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Characteristics</td>
<td>( k )</td>
<td>Effect sizes, n</td>
<td>( B_0 ) (intercept), ( g ) (95% CI)</td>
<td>( B_1 ) (slope), ( g ) (95% CI)</td>
<td>( F ) (df(_1), df(_2))</td>
<td>( P ) value</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>--------</td>
<td>-----------------</td>
<td>----------------------------------------</td>
<td>----------------------------------------</td>
<td>-------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Cognitive behavioral and mindfulness</td>
<td>6</td>
<td>46</td>
<td>0.35 (0.12 to 0.58) **</td>
<td>0.05 (−0.02 to 0.29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other or multiple (eg, motivational)</td>
<td>20</td>
<td>171</td>
<td>0.16 (0.04 to 0.29) **</td>
<td>−0.14 (−0.30 to 0.01)(^{\text{d}})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Athoretical or not specified</td>
<td>3</td>
<td>5</td>
<td>0.41 (−0.04 to 0.85)(^{\text{e}})</td>
<td>0.10 (−0.35 to 0.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intervention technological features</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personalization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.37 (1, 706)</td>
<td>.24</td>
</tr>
<tr>
<td>Absent (RC)</td>
<td>63</td>
<td>575</td>
<td>0.28 (0.21 to 0.35) ***</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>17</td>
<td>133</td>
<td>0.19 (0.07 to 0.32) **</td>
<td>−0.09 (−0.23 to 0.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tailoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.84 (1, 706)</td>
<td>.36</td>
</tr>
<tr>
<td>Absent (RC)</td>
<td>45</td>
<td>399</td>
<td>0.28 (0.20 to 0.37) ***</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>34</td>
<td>309</td>
<td>0.23 (0.13 to 0.32) ***</td>
<td>−0.06 (−0.18 to 0.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.17 (1, 706)</td>
<td>.68</td>
</tr>
<tr>
<td>Absent (RC)</td>
<td>70</td>
<td>628</td>
<td>0.26 (0.20 to 0.33) ***</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>9</td>
<td>80</td>
<td>0.22 (0.05 to 0.40) *</td>
<td>−0.04 (−0.23 to 0.15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gamification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.78 (1, 706)</td>
<td>.38</td>
</tr>
<tr>
<td>Absent (RC)</td>
<td>60</td>
<td>531</td>
<td>0.24 (0.17 to 0.31) ***</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>19</td>
<td>177</td>
<td>0.31 (0.18 to 0.44) ***</td>
<td>0.07 (−0.08 to 0.21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intervention support features</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orientation to or training on mTDI(^{h})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.09 (1, 678)</td>
<td>.77</td>
</tr>
<tr>
<td>Absent (RC)</td>
<td>48</td>
<td>361</td>
<td>0.27 (0.19 to 0.35) ***</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>29</td>
<td>319</td>
<td>0.25 (0.15 to 0.35) ***</td>
<td>−0.02 (−0.15 to 0.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other in-person element</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.004 (1, 707)</td>
<td>.95</td>
</tr>
<tr>
<td>Absent (RC)</td>
<td>68</td>
<td>637</td>
<td>0.27 (0.20 to 0.33) ***</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>12</td>
<td>72</td>
<td>0.27 (0.11 to 0.43) ***</td>
<td>0.006 (−0.16 to 0.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reminders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.91 (1, 696)</td>
<td>.17</td>
</tr>
<tr>
<td>Absent (RC)</td>
<td>40</td>
<td>233</td>
<td>0.31 (0.21 to 0.40) ***</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>37</td>
<td>465</td>
<td>0.22 (0.13 to 0.30) ***</td>
<td>−0.09 (−0.22 to 0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guidance, coaching, and feedback</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.24 (1, 707)</td>
<td>.62</td>
</tr>
<tr>
<td>Absent (RC)</td>
<td>59</td>
<td>531</td>
<td>0.28 (0.20 to 0.35) ***</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>21</td>
<td>178</td>
<td>0.24 (0.12 to 0.36) ***</td>
<td>−0.04 (−0.18 to 0.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supportive accountability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.14 (1, 707)</td>
<td>.71</td>
</tr>
<tr>
<td>Absent (RC)</td>
<td>62</td>
<td>538</td>
<td>0.27 (0.20 to 0.35) ***</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>18</td>
<td>171</td>
<td>0.25 (0.12 to 0.38) ***</td>
<td>−0.03 (−0.18 to 0.12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervised skills practice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.62 (1, 707)</td>
<td>.43</td>
</tr>
<tr>
<td>Absent (RC)</td>
<td>72</td>
<td>651</td>
<td>0.26 (0.19 to 0.33) ***</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>8</td>
<td>58</td>
<td>0.34 (0.14 to 0.54) ***</td>
<td>0.08 (−0.13 to 0.30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dosage: prescribed frequency of use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.39 (4, 704)</td>
<td>.002(^{c})</td>
</tr>
</tbody>
</table>

\(^{h}\)Dosage: prescribed frequency of use
Given this moderation effect, we also ran all analyses with only the higher-quality studies—that is, studies that achieved a total quality index >20, which denotes studies that, on average, surpassed benchmarks for average-quality research methods (yielding 428 ESs across posttest and follow-up assessments) was $g=0.20$ ($P<.001$; 95% CI 0.13-0.27). A funnel plot analysis revealed that publication bias was unlikely, with no studies missing on the left side of the funnel plot (Figure 2). Indeed, the trim-and-fill algorithm suggested that, if anything, 78 ESs from 34 studies were missing at the right side of the funnel plot, suggesting a possible selection bias that excluded studies with larger ESs. After imputation of these missing ESs, an adjusted overall effect was estimated, which produced an average ES of $g=0.40$ ($P<.001$; 95% CI 0.33-0.46), somewhat larger than our initially estimated overall effect ($g=0.13$). Nevertheless, it is important to note that the trim-and-fill analysis does not take the dependency in ESs into account. An Egger regression test, which better models these dependencies among ESs, revealed that SE was a significant positive predictor of ESs ($\beta_3=1.65$, $P<.001$; 95% CI 1.07-2.23), which may indicate publication bias rather than selection bias.

Table 2 (full set of results available from authors upon request). The average ES across all possible comparisons within the 38 higher-quality studies (yielding 428 ESs across posttest and follow-up assessments) was $g=0.20$ ($P<.001$; 95% CI 0.13-0.27). A funnel plot analysis revealed that publication bias was unlikely, with no studies missing on the left side of the funnel plot (Figure 2). Indeed, the trim-and-fill algorithm suggested that, if anything, 78 ESs from 34 studies were missing at the right side of the funnel plot, suggesting a possible selection bias that excluded studies with larger ESs. After imputation of these missing ESs, an adjusted overall effect was estimated, which produced an average ES of $g=0.40$ ($P<.001$; 95% CI 0.33-0.46), somewhat larger than our initially estimated overall effect ($g=0.13$). Nevertheless, it is important to note that the trim-and-fill analysis does not take the dependency in ESs into account. An Egger regression test, which better models these dependencies among ESs, revealed that SE was a significant positive predictor of ESs ($\beta_3=1.65$, $P<.001$; 95% CI 1.07-2.23), which may indicate publication bias rather than selection bias.

### Differences in Effects of mTDIs Based on Methodological Characteristics

#### Study Quality and Publication Bias

Overall study quality significantly moderated the overall effect of mTDIs in such a way that ESs decreased as the quality index increased (Table 1). The slope indicated that for every 1-point increase in quality index score, the effect decreased by 0.02. Given this moderation effect, we also ran all analyses with only the higher-quality studies—that is, studies that achieved a total quality index >20, which denotes studies that, on average, surpassed benchmarks for average-quality research methods [89]. Unless otherwise noted in the relevant presentation of results in the following sections, the pattern and significance of the results with this reduced, higher-quality sample of studies were identical to those of the full sample of studies. However, in cases where the statistical significance of results shifted when tested with only higher-quality studies, the results from analyses with only the higher-quality studies are presented separately in

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>$k^b$</th>
<th>Effect sizes, n</th>
<th>$B_0$ (intercept), $g$ (95% CI)</th>
<th>$B_1$ (slope), $g$ (95% CI)</th>
<th>$F$ (df$_1$, df$_2$)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dosage: prescribed duration of intervention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weeks</td>
<td>78</td>
<td>697</td>
<td>0.27 (0.21 to 0.34)***</td>
<td>−0.003 (−0.01 to 0.01)</td>
<td>0.35 (1, 695)</td>
<td>.56</td>
</tr>
<tr>
<td>Sessions</td>
<td>54</td>
<td>414</td>
<td>0.33 (0.24 to 0.41)***</td>
<td>−0.0001 (−0.0003 to 0.0002)</td>
<td>0.44 (1, 412)</td>
<td>.51</td>
</tr>
</tbody>
</table>

---

The right columns list the omnibus $F$ test and $P$ value for each moderation test. The middle columns list the intercept ($B_0$), or mean effect size, and slope ($B_1$), an estimated unstandardized regression coefficient, of the relevant Hedges $g$ statistics, with CIs around each. Effects and slopes that differ significantly from 0 are denoted with asterisks in the intercept ($B_0$) and slope ($B_1$) columns, respectively. For categorical moderators, each intercept represents the mean effect of a category, whereas each slope represents the difference in the mean effect between the category and reference category. Depending on its sign, the slope of a continuous moderator represents an increase or decrease in the effect size with each unit increase in the variable.

Number of studies with relevant effect size data for a given row. In cases of multiple interventions or comparisons, some studies were counted in multiple rows; thus, these numbers sometimes exceeded 80. Owing to missing data, some counts fall short of 80. Further details on what was included in different categories of the included moderators are provided in the Methods section.

Significance of moderation analysis changed when conducted on a subsample of the highest-quality studies ($k=38$; Table 2).

RC: reference category.

VR: virtual reality.

mTDI: mobile technology–delivered intervention.

$P<.10$.

$P<.05$.

$P<.01$.

$P<.001$.

---

aThe right columns list the omnibus $F$ test and $P$ value for each moderation test. The middle columns list the intercept ($B_0$), or mean effect size, and slope ($B_1$), an estimated unstandardized regression coefficient, of the relevant Hedges $g$ statistics, with CIs around each. Effects and slopes that differ significantly from 0 are denoted with asterisks in the intercept ($B_0$) and slope ($B_1$) columns, respectively. For categorical moderators, each intercept represents the mean effect of a category, whereas each slope represents the difference in the mean effect between the category and reference category. Depending on its sign, the slope of a continuous moderator represents an increase or decrease in the effect size with each unit increase in the variable.

bNumber of studies with relevant effect size data for a given row. In cases of multiple interventions or comparisons, some studies were counted in multiple rows; thus, these numbers sometimes exceeded 80. Owing to missing data, some counts fall short of 80. Further details on what was included in different categories of the included moderators are provided in the Methods section.

cSignificance of moderation analysis changed when conducted on a subsample of the highest-quality studies ($k=38$; Table 2).

d RC: reference category.

e N/A: not applicable (as the slope represents a comparison with the reference category).

f VR: virtual reality.

g $P<.10$.

h mTDI: mobile technology–delivered intervention.

* $P<.10$.

** $P<.05$.

*** $P<.01$.

Table 2 (full set of results available from authors upon request).
Table 2. Moderators of the effectiveness of mobile technology–delivered interventions for youth: higher-quality studies onlya.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>kᵇ</th>
<th>Effect sizes, n</th>
<th>B₀ (intercept), g (95% CI)</th>
<th>B₁ (slope), g (95% CI)</th>
<th>F test (df₁, df₂)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Methodological characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study quality</td>
<td>38</td>
<td>428</td>
<td>0.20 (0.11 to 0.29)***</td>
<td>-0.0005 (–0.03 to 0.03)</td>
<td>0.001 (1, 426)</td>
<td>.97</td>
</tr>
<tr>
<td><strong>Comparison group type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No intervention (RCc; eg, wait-list)</td>
<td>21</td>
<td>215</td>
<td>0.26 (0.18 to 0.35)***</td>
<td>N/Ad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inert (eg, placebo or information-only)</td>
<td>15</td>
<td>112</td>
<td>0.14 (0.05 to 0.23)**</td>
<td>–0.13 (–0.23 to –0.03)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical (eg, established intervention)</td>
<td>7</td>
<td>101</td>
<td>0.12 (–0.01 to 0.26)e</td>
<td>–0.14 (–0.29 to 0.01)e</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Youth Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean age (years)</td>
<td>36</td>
<td>407</td>
<td>0.19 (0.11 to 0.26)***</td>
<td>0.01 (–0.0002 to 0.03)e</td>
<td>3.75 (1, 405)</td>
<td>.05</td>
</tr>
<tr>
<td><strong>Risk level and type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General sample not selected for risk (RC)</td>
<td>17</td>
<td>175</td>
<td>0.13 (0.03 to 0.24)**</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonmental health (ie, medical) risks</td>
<td>1</td>
<td>2</td>
<td>0.11 (–0.50 to 0.72)</td>
<td>–0.03 (–0.64 to 0.59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological or mental health at-risk sample</td>
<td>16</td>
<td>229</td>
<td>0.25 (0.15 to 0.36)**</td>
<td>0.12 (–0.03 to 0.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological clinical sample (diagnosis)</td>
<td>4</td>
<td>22</td>
<td>0.31 (0.07 to 0.55)*</td>
<td>0.18 (–0.09 to 0.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intervention characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prescribed frequency of use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>As much as feasible; ≥4 days per week (RC)</td>
<td>22</td>
<td>306</td>
<td>0.23 (0.14 to 0.32)**</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some days, or more than once a week</td>
<td>5</td>
<td>27</td>
<td>0.22 (0.02 to 0.42)*</td>
<td>–0.01 (–0.22 to 0.21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>About once a week</td>
<td>3</td>
<td>36</td>
<td>0.40 (0.15 to 0.66)**</td>
<td>0.18 (–0.09 to 0.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-time session</td>
<td>2</td>
<td>15</td>
<td>0.13 (–0.20 to 0.46)</td>
<td>–0.10 (–0.44 to 0.25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not stated, when needed, or at user discretion</td>
<td>6</td>
<td>44</td>
<td>0.03 (–0.14 to 0.20)</td>
<td>–0.20 (–0.39 to –0.01)*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aThis table presents moderation results for higher-quality studies (k=38) only in cases where the statistical significance of the moderation effect differs from the full-sample (k=80) results presented in Table 1. The right columns list the omnibus F test and P value for each moderation test. The middle columns list the intercept (B₀), or mean effect size, and slope (B₁), an estimated unstandardized regression coefficient, of the relevant Hedges g statistics, with CIs around each. Effects and slopes that differ significantly from 0 are denoted with asterisks in the intercept (B₀) and slope (B₁) columns, respectively. For categorical moderators, each intercept represents the mean effect of a category, whereas each slope represents the difference in the mean effect between the category and reference category. Depending on its sign, the slope of a continuous moderator represents an increase or decrease in the effect size with each unit increase in the variable.
bNumber of studies with relevant ES data for a given row. In cases of multiple interventions or comparisons, some studies were counted in multiple rows; thus, these numbers sometimes exceeded 38. Owing to missing data, some counts fell short of 38. Further details on what was included in different categories of the included moderators are provided in the Methods section.
cRC: reference category.
dN/A: not applicable (as the slope represents a comparison with the reference category).
eP<.10.
*P<.05.
**P<.01.
***P<.001.
Timing of Outcome Assessment
There were no significant differences in ESs immediately after the intervention versus those at longer-term follow-up assessments (Table 1).

Outcome Type
The effectiveness of mTDIs varied as a function of the type of youth outcome that was targeted or assessed. The results in Table 1 indicate that there were statistically significant, positive effects of mTDIs on all of the coded outcome categories: general psychological distress or well-being, internalizing distress, noninternalizing mental health concerns, psychosocial strategies and skills, health-related outcomes, and other outcomes (see the Methods section). However, the other outcomes showed significantly lower ESs than the reference category, on average.

Comparison Group Type
Contrary to expectations, in the full sample, the comparison group type did not moderate ES, such that effects were not statistically different across studies using no-intervention (eg, wait-list) or inert (eg, placebo or information-only) comparison groups, as well as studies using clinical treatments as their comparison group (Table 1). However, among the higher-quality studies, the results were more in line with our hypotheses, in that studies using inert comparison groups produced lower ESs than studies using no-intervention control groups, and studies using a clinical comparison no longer showed statistically significant effects on youth outcomes (Table 2).

Differences in Effects of mTDIs Based on Youth Characteristics
The results showed that the mean age of the youth participants did not moderate the impact of the mTDIs (Table 1). Among the higher-quality studies, there was an effect right at the $P=0.05$ threshold, such that the older the mean age of participants, the stronger the effect (Table 2). There were no differences in the study ESs as a function of the youth gender breakdown in the sample. Missing data on race and ethnicity limited our ability to analyze this variable as a moderator.

Youth level and type of risk significantly moderated intervention effects in the full sample (Table 1), such that samples with nonmental health (ie, medical) risks showed larger effects of mTDIs than general, unselected youth samples. However, in the subsample of higher-quality studies, this moderating effect was not found; in fact, the medical risk category dropped to one study and was no longer significantly different from zero (Table 2).

Differences in Effects of mTDIs Based on Intervention Characteristics
Primary Type of Technology
Moderation analysis did not detect statistically significant differences in the impact of mTDIs based on the primary type of technology: mTDIs were effective—and similar in their impact on youth outcomes—whether delivered on a smartphone or tablet, a presmartphone mobile device, or a mobile VR or handheld video game (Table 1).

Guiding Theoretical Framework
As hypothesized, both cognitive behavioral and mindfulness-or acceptance-based interventions (as well as interventions that blended these 2 orientations) had significant effects on youth outcomes. Interventions grounded in one or multiple other theoretical frameworks also yielded significant effects and did not appear to differ systematically in their effects from cognitive behavioral interventions. Those mTDIs that were atheoretical or did not specify a guiding theoretical framework did not significantly differ from zero in their impact on youth outcomes, and the CI around their intercept (mean effect) was quite wide, indicating considerable heterogeneity. Of note, these studies were rare ($k=3$), and all 3 studies were dropped from the analysis of higher-quality studies; however, the overall pattern of results remained the same.
Technological and Support Features

Exploratory analyses of the impact of intervention features and support failed to detect significant moderation of intervention effects based on the presence or absence of various technological features of the mTDI, including personalization, tailoring, social components, or gamification elements. Similarly, there were no differences in effects for mTDIs that integrated various support features, such as a training or orientation to the mTDI; some other in-person element; reminders to use the app; human or bot guidance, coaching, feedback in mTDI use; provision of supportive accountability; or supervised practice of skills taught by the mTDI. Although no significant differences were found between the absence and presence of any of these features and support types, it is notable that mTDIs both with and without each of these features had significant and positive mean effects (Table 1).

Dosage: Prescribed Frequency and Duration of Use

There was a significant moderation effect for the prescribed frequency of mTDI use (Table 1). All prescribed use frequencies, except for leaving use to user discretion (including unstated use prescriptions), had a statistically significant impact on youth outcomes. Those mTDIs that involved prescribed use about once per week or were a single session yielded higher ESs than the reference category, which involved prescriptions of more frequent mTDI use (ie, at least 4 days per week or much as feasible). However, this effect was not retained in the sample of higher-quality studies: The 2 remaining studies that prescribed a one-time session no longer yielded ESs that differed from zero statistically, and the prescribed use frequency was no longer a significant moderator of ESs (Table 2). Additional moderation analyses probing the number of prescribed weeks or sessions of mTDI use did not detect significant effects for the prescribed duration of the intervention, whether the number of intervention sessions or weeks (Table 1). Although we also intended to examine the moderating effect of completed dosage (frequency, weeks, and sessions), there were substantial missing data, precluding meaningful analysis of these moderators.

Discussion

Principal Findings and Comparisons With Prior Work

To our knowledge, this study represents the first review and meta-analysis of mTDIs for a wide variety of youth well-being outcomes, an area of research that has grown rapidly in the past decade. Rigorous searches of the published and unpublished literature in this area yielded 80 studies evaluating 83 mTDIs for youth. A 3-level meta-analysis revealed an overall Hedges g of 0.27 across all youth outcomes and follow-up assessments, indicating a small effect that is generally consistent with the observed impact of mTDIs in other meta-analyses [42,62,67-69]. This finding addresses a critical gap in the existing literature in that most previous meta-analyses have focused solely on the effects of mTDIs in adult populations [42,53,63], and the few studies focusing specifically on youth have aggregated across diverse types of mobile and nonmobile technologies [64,65] or limited their scope to a specific subset of youth disorders (eg, internalizing disorders [67]).

It is worth noting that our sample included many studies of mTDIs that were still relatively early in their development and were, therefore, primarily interested in evaluating the feasibility and acceptability of the technology, although they also included measures of more distal mental health outcomes that they ultimately aimed to influence. Therefore, our analyses may underestimate, to some extent, the impact that these mTDIs would have had in larger or longer efficacy trials more specifically designed to influence youth mental health outcomes. As this literature continues to mature, it will be important to focus the inclusion criteria more specifically on studies that measure the effects of mTDIs on more distal mental health outcomes as their primary focus.

Our publication bias analyses yielded conflicting findings that were difficult to interpret, given the lack of conventions for analyzing publication bias in 3-level models. However, it is worth noting that despite rigorous screening criteria for the methodology of included studies, our coding revealed significant variability in study quality, with less than half of the studies in our sample comprising effects that surpassed our defined standards for typical research practices [89]. Study quality appeared to significantly influence the observed ESs such that as the score for study quality increased, the ESs generally decreased. Only a few prior meta-analyses of TDIs have directly assessed the influence of study quality on ES and found no impact [57,201]. However, our findings are consistent with some previous findings linking study quality to observed ESs for other psychological interventions [202,203] and suggest that attention to rigorous experimental methods, such as the reporting of intent-to-treat analyses and the use of well-validated and reliable assessment tools, are essential to accurately identify the impact of mTDIs for youth.

Methodological Characteristics Affecting Outcomes

Given the substantial heterogeneity across ESs, both within and between studies, we explored several moderators as predictors of this variability. Interestingly, the ESs were similar in the immediate posttest assessments and longer-term follow-up assessments. This was true despite the fact that follow-up assessments occurred, on average, at 11.52 weeks, and ranged in length up to 43.14 weeks, after the active intervention period concluded. This finding is in contrast to the decrease in the effectiveness over time of some in-person mental health treatments [204-206] and suggests that the impact of mTDIs endures over time. It is possible that these enduring effects are because mTDIs are more easily integrated into youths’ lives, therefore leading to either greater generalizability of the intervention effects, more lasting engagement with the mTDI, or both.

Somewhat contrary to expectations, moderator analyses in the overall sample also revealed that mTDIs had a similar impact on youth outcomes regardless of whether they were compared with a no-intervention control, such as a wait-list, or a more active comparison group, such as an information-only condition or usual clinical care. This helps to rule out the effects of expectancies, demand characteristics, or nonspecific effects...
accounting for the benefits of mTDIs in youth. This finding contributes to the somewhat mixed literature on this topic, with some past reviews of TDIs with both youth and adult samples finding that, more generally, ESs tend to differ based on comparison type (eg, higher for wait-list vs more active comparisons) [45,52,58,69,70], whereas others indicate that TDIs tend to be similarly effective across various types of study designs (eg, the study by Farrer et al [85]). Indeed, even in this meta-analysis, some findings shifted when only higher-quality studies were analyzed, such that studies with information-only or placebo comparisons yielded lower ESs than studies with no-intervention control groups, and studies with clinical comparison groups no longer showed statistically significant effects. Moreover, it is worth noting that the specific nature of the comparison group varied quite widely across studies, even within a particular coded category. As such, future research should continue to explore the marginal benefits of mTDIs over other available interventions.

The positive impact of mTDIs was observed across the diverse youth outcome categories assessed by each study, with the largest ESs for psychosocial strategies and skills (eg, emotional self-awareness, self-efficacy, and coping) and internalizing symptoms such as depression and anxiety, followed by general psychological distress and well-being, health concerns and health-related behaviors, and other noninternalizing mental health concerns (eg, attention difficulties, aggression, or delinquency). The smallest ES was observed for other outcomes (eg, knowledge, peer relationship quality, and stereotype threat), which showed significantly smaller effects than the reference category of general psychological distress. However, it is difficult to interpret this finding, given that our other category contained a diverse set of outcomes, many of which were coded very infrequently. As such, the overall ES for this other category is not necessarily reflective of the lower impact of mTDIs on each of these less commonly coded outcome types, and future research should continue to explore the scope of the impact of mTDIs on diverse youth problems. Nevertheless, these findings regarding outcome types generally suggest that mTDIs can be effective in treating a wide array of problems across youth development, including diverse areas of psychopathology (ie, both internalizing and externalizing domains), in addition to a number of cognitive, behavioral, and social risk factors that are often associated with poor mental health. This is a significant contribution to the literature, which has previously focused on narrower sectors of outcomes when analyzing the effectiveness of TDIs and mTDIs for youth [51,67,68].

Youth Characteristics Affecting Outcomes

In the overall sample, there was no association between average youth age and the impact of mTDIs. However, when lower-quality studies were excluded from the analysis, the effects of youth age emerged more strongly, with ESs increasing as the mean participant age increased. This effect was right at the threshold for statistical significance (P=.05) and should, therefore, be interpreted with caution and replicated in future studies; however, this finding that suggests stronger effects of mTDIs for older youth is consistent with some past literature on TDIs more generally [51,54,55,65,72]. As many mental health interventions, including TDIs, were originally developed with adults in mind and only later adapted for youth at various stages of development, it is perhaps not surprising that mTDIs could have a more robust impact on older adolescents and young adults. For example, these youth may have greater internal motivation to engage with the intervention and be better able to interact with and adhere to the cognitive or behavioral skills taught by the mTDIs. However, given the complex ways in which developmental stages interact with risk for diverse mental health problems, as well as the effectiveness of mental health interventions, future research should continue to probe interactions between youth age and other dimensions of mTDIs (eg, level of human support, guiding theoretical framework, and availability of a social component) in predicting the impact of mTDIs.

Youth risk characteristics significantly moderated intervention effects in the full sample, with studies in which youth were selected for indicators of medical risk (eg, youth diagnosed with spina bifida or about to undergo surgery or another medical procedure) showing an average ES more than double that of studies with general, unselected samples of youth. Samples of youth with psychological risk (either clinically significant or subclinical risk) fell somewhere in the middle. However, in the analyses that excluded the lowest-quality studies, this moderating effect was no longer observed. The lack of differential findings for the impact of mTDIs on outcomes for youth with clinically significant versus subclinical risk, even when compared with general or unselected samples, is somewhat consistent with previous findings on TDIs, which tend to be quite mixed in terms of the impact of TDIs for youth with a variety of risk profiles [47,53,60,65]. Future research should continue to explore the presenting problems and risk indicators that are the best fit for referral to mTDIs versus more or less intensive interventions.

Other youth characteristics, in addition to age and risk factors, were not demonstrated to predict observed ESs. Youth race and ethnicity were reported inconsistently and according to widely varying conventions and, therefore, could not be tested as moderators of ESs. Furthermore, these identities and lived experiences are intertwined with structural inequalities and systems of discrimination and oppression that are more important for research to assess and relate to well-being outcomes. Consistent with several previous reviews [42,59], the breakdown of youth gender in the study sample did not predict ESs, although it is worth noting that studies infrequently made a note of nonbinary gender categories. These findings point to the need for a more careful and nuanced assessment of youth identities and lived experiences, including those connected to race, cultural identity, gender, sexual identity, and socioeconomic background, in studies testing the impact of mTDIs on youth.

Intervention Characteristics Affecting Outcomes

Smartphone- or tablet-based mTDIs were by far the most commonly reported primary technology in our sample of studies relative to presmartphone mobile devices or other (mobile VR and handheld video game) technologies. The type of technology did not appear to moderate ESs, with each of these types of mTDIs yielding average ESs that were statistically significant.
and of a similar size. Although our sample was limited in number, mobile VR technologies are promising avenues for further research, especially given the effectiveness of these technologies for conditions such as posttraumatic stress disorder, depression, and pediatric pain and anxiety during medical procedures [207,208]. Although no studies in our sample used wearable devices as the primary type of technology, a handful used a smartphone along with some sort of wearable biosensor such as a sleep monitor or a physical activity wristwatch [113,183,188,195]. As these technologies are likely to become more common over time, research should continue to explore their effectiveness as a primary or supplemental feature of mTDIs for youth well-being.

Consistent with the previous literature on TDI and mTDIs for youth and adults [51,53,54,69], both cognitive behavioral and mindfulness- or acceptance-based interventions (as well as interventions that blended these 2 orientations) had significant effects on youth outcomes. In traditional in-person treatment settings, cognitive behavioral interventions, including third-wave cognitive behavioral treatments that include components of mindfulness and acceptance, have become increasingly popular as empirical support has grown for their effectiveness in treating a wide range of childhood disorders, including anxiety, depression, conduct or aggression problems, and attention difficulties [209]. However, a growing body of literature shows that many youths and families are not able to access these gold standard evidence-based treatments, whether because of lack of availability in their community or issues with accessing mental health care in general, such as cost and stigma [26,210,211]. Therefore, it is encouraging to see that the effectiveness of these evidence-based interventions can be translated into low-cost, mobile technology–delivered formats that can reach far larger numbers of youth, and perhaps in a way that is more generalizable to the naturalistic environments of their lives. Interventions in our sample that were grounded in one or multiple other theoretical frameworks, such as positive psychology or motivational interviewing, also yielded significant effects and did not appear to differ systematically in their effects from cognitive behavioral interventions. Although few studies have examined (m)TDIs using these theoretical approaches, these findings are consistent with previous research that has evaluated the impact of these specific theoretical orientations [202,212]. Our sample also included 3 mTDIs that did not specify a guiding theoretical framework [154,166,171], and collectively, they did not significantly differ from zero in their impact on youth outcomes. These findings should be interpreted with caution, given the small number of studies, wide CIs around their intercepts (mean effects), and the fact that these studies were excluded from the analysis of higher-quality studies. Future research should continue to explore the impact of mTDIs grounded in diverse theoretical frameworks. With the vast and rapidly growing number of available mTDIs purporting to support the well-being and mental health of youth, it is critical to ascertain the theoretical frameworks that may lend themselves best to developing active interventions with strong empirical support for their effectiveness in a mobile technology–delivered format.

As mTDIs have become increasingly popular, many have begun to incorporate additional technological features intended to better leverage the technology-based format to engage and sustain users’ attention. For example, some apps may personalize features of the intervention to the user’s personal preferences or tailor the intervention based on a user’s in-the-moment responses [75]. Others may include a social component, such as integration with social media platforms or a chat forum, or incorporate aspects of gamification, such as challenges or quests associated with points or badges [213]. Our moderator analyses showed no influence of these features on the ESs. However, it should be noted that many of these features are still relatively uncommon in mTDIs tested by research. For example, only 12% (10/83) of mTDIs in our meta-analysis mentioned a social component, and only 22% (18/83) described elements of personalization. Thus, the importance of these features may become more apparent as mTDIs targeting youth well-being begin to incorporate them more regularly and with greater proficiency. It is also likely that the most impactful mTDIs use an effective combination of these features to engage youth rather than simply incorporating one or another single design feature.

Given extensive theories [77,214,215] and some prior research on the benefits of outside guidance on engagement with mTDIs [53,57,216], we were somewhat surprised to see that the incorporation of support features, such as the provision of supportive accountability for technology use or supervised practice of the skills introduced by the mTDI, did not significantly moderate the ESs for mTDIs for youth. However, research in this area has been quite mixed, with several other studies finding little or no benefit from the inclusion of coaching or human support [42,54,55]. It is possible that mTDIs that do not rely on any component of human or bot support tend to be designed in a more comprehensive and self-contained way to offset this lack [42]. Moreover, the lack of significant moderation findings for these features in our meta-analysis does not necessarily indicate that these features are unimportant to the success of mTDIs in youth. Our meta-analysis captured an unusual sample of mTDI users, given that all effects were evaluated within the context of researcher-guided studies. As such, all participants were likely exposed to greater-than-usual accountability and support for their technology use as a function of taking part in a research study. This level of baseline accountability may have made it difficult to observe the added benefits of other forms of guidance or support. In addition, there are likely several kinds of informal human support—such as guidance or support from caregivers, teachers, or peers—that influence youth but were not typically assessed or reported in our sample of studies. Future research should continue to explore the kinds of support that are needed to maximize engagement with mTDIs for youth mental health and the ways in which these supports interact with factors such as youth age or risk characteristics (eg, younger or more clinically at-risk youth may require greater support).

Finally, moderation analyses provided tentative evidence that mTDIs can be effective regardless of the prescribed frequency or duration of use. In the overall sample, the number of prescribed weeks or sessions of use did not moderate the ESs.
Moreover, mTDIs yielded significant ESs across all prescribed use frequencies, except for mTDIs that did not prescribe a use frequency or left use up to the user’s discretion. In the higher-quality sample, single-session interventions no longer yielded a significant ES; however, this finding should be interpreted with caution as only 2 single-session interventions remained in the higher-quality analytic sample.

These findings add to a growing body of mixed findings regarding the impact of prescribed and actual mTDI use on intervention outcomes [57,85,86]. Reviews of technology-based mental health interventions often highlight significant problems with treatment initiation and dropout [217-219], particularly for self-guided treatments that involve lower levels of structure and prescriptive guidance [220,221]. As such, it is critical for meta-analytic research to continue to explore trends in whether and how the prescribed dosage of mTDIs influences youth outcomes, with a particular focus on how different types of prescriptive guidance fit best with users’ specific needs. For example, youth with more severe clinical diagnoses may require a different dosage of mTDI than those engaging with a prevention-oriented mTDI designed to improve general well-being. Notably, there was wide variability in how studies reported on mTDI dosage and adherence. We chose to analyze the prescribed dosage and frequency of mTDI use, given that these statistics were most consistently reported. It would be ideal to analyze the actual completed use or uptake of mTDIs among participants as well; however, this was reported inconsistently among the studies. Several studies reported participant use only among study completers or dropped unengaged or low-engaged users from the analysis [121,127,152], whereas others used financial incentives for protocol compliance [151], potentially introducing bias into the use statistics recorded in research studies that have additional levels of accountability built into the protocol.

Limitations and Future Directions

The results of this meta-analysis should be interpreted with several limitations in mind. First, the quality of any meta-analysis is limited by the quality of the available primary studies. Only 60% of our coded ESs, which came from 48% (38/80) of the included studies, met our high-quality standards. We attempted to address this issue by running all analyses on both the full sample of studies and a subsample of higher-quality studies. Nevertheless, as a substantial number of studies and ESs of lower quality were dropped from the higher-quality subsample analysis, the statistical power declined, and some moderator categories could not be examined. Future research in this area should attend to existing procedures for designing and reporting on high-quality clinical intervention research. For example, there is a need for more studies that use larger sample sizes and retain larger percentages of their participants (regardless of their mTDI engagement), include more reliable outcome measurements, and use intent-to-treat analyses, as well as studies conducted by authors who were not involved with app development and are, therefore, able to provide a more unbiased assessment of the mTDI’s impact.

Relatedly, our coding scheme yielded incomplete data for many of our hypothesized moderators, given the variability in reports on characteristics of the tested mTDI (eg, human support features, duration and frequency of use), as well as youth characteristics (eg, race and ethnicity, gender, and risk characteristics). Therefore, additional studies that carefully document these kinds of data are needed to more thoroughly test the various moderators of the overall effects of mTDIs.

Given the lack of standard approaches for assessing publication bias within a 3-level meta-analysis, we applied 2 different techniques that produced conflicting findings. The trim-and-fill analysis pointed toward a potential underestimation of the true overall effect, whereas the Egger test pointed toward a potential overestimation of the true overall effect (and thus publication bias). The Egger test—in which ES dependencies are modeled—is likely more valid than the trim-and-fill results for our multilevel study. Nevertheless, the results of both techniques should be interpreted with caution, as neither was developed for a 3-level meta-analysis, and both rely on an assumption of homogeneity in ESs, which is often not met in meta-analytic studies, including this study.

As noted previously, another limitation in interpreting the present findings is that youth mTDI use likely differs within versus outside of a research study. At the very least, research participants tend to be much more informed about intervention goals and receive more structured support in the process of using an mTDI than users outside the research context. In more naturalistic settings, such as clinical practice or completely self-guided use, mTDIs are likely to be used much more flexibly and may be adapted to the needs and circumstances of individual youths.

When considering the implications of the present findings for the use of mTDIs in clinical and other naturalistic settings, it is also essential to consider issues such as accessibility and cultural sensitivity of mTDIs, which were not examined in this meta-analysis. Some types of mTDI technologies, such as VR headsets, may be prohibitively expensive and less readily accessible to some youth and families, particularly those living or seeking treatment in low-resource settings. Moreover, wide variability in families’ digital literacy and cultural norms around using technology to improve well-being is likely to play a key role in the effectiveness of these types of interventions. The studies in this meta-analysis were largely limited to Western cultural contexts, with very little research emerging from certain parts of the world (eg, our search did not yield any research from the African continent). In addition, reporting on youth characteristics such as socioeconomic status, race, and ethnicity was quite limited and followed widely varying conventions. Taken together, these issues limited our ability to test questions related to the cultural responsivity or tailoring of particular mTDIs based on youth cultural backgrounds—important questions that future research will need to explore to fulfill the promise of mTDIs for youth living in communities traditionally underserved by available in-person prevention and intervention programs.

In addition, participant ratings of social validity (ie, acceptability of mTDIs and user satisfaction), among the 25% (20/80) of studies for which these data were available, averaged under 60%, indicating that these technologies likely have room for...
improvement in user interface and user experience design. Further research should more diligently assess for different aspects of social validity, including qualitative feedback, and relate these elements to outcomes and potential moderators (eg, age of users, in-app features, and clinical severity of users), with the ultimate goal of improving the engagement and uptake, and thus impact, of mTDIs for youth.

Finally, future research should continue to explore the ideal setting and level of support for various mTDIs. The optimal approaches to integrating mTDIs with other mental health tools, as well as face-to-face interaction with mental health providers, remain largely an open question at this time. Although in some cases, mTDIs may serve as low-cost and accessible substitutions or adjunctive supports for face-to-face intervention programs in areas with limited access to health professionals, mTDIs may also be valuable as a way of socializing some youth to psychosocial interventions, with the goal of eventually connecting families to more traditional face-to-face services. In other cases, mTDIs may be most useful when accompanied by the support of a clinician or paraprofessional, such as a teacher, mentor, or academic advisor who guides the youth through the technology-based intervention [118,120,158].

Study Strengths and Conclusions
This is the first comprehensive 3-level meta-analysis to evaluate the effects of mTDIs on diverse aspects of well-being in youth. We built on prior work by taking an inclusive but rigorous approach to testing the impact of interventions using various types of mobile technologies on diverse outcomes across childhood, adolescence, and young adulthood. Using a 3-level approach to meta-analysis, we were able to synthesize all relevant ESs while accounting for both within- and between-study heterogeneity in ESs and maximizing the statistical power of the analyses. Moreover, we coded a comprehensive set of more than 3 dozen potential moderators of study effects and found sufficient information to analyze the moderating role of >20 of these variables, including technological and support features (eg, human support and availability of in-app reminders) that are hypothesized to be critical to the success of these interventions but have rarely been tested as predictors of effects in previous meta-analyses.

Our synthesis of primary research confirms the significant benefits of mTDIs across a variety of psychosocial outcomes, comparison types (ie, no intervention, inert, and clinical), and time points (both immediate postintervention and longer-term follow-up effects). Although additional high-quality research on which kinds of mTDIs are most effective and under what conditions is clearly needed, we conclude that mTDIs have the potential to improve multiple aspects of youth well-being, and may confer significant, durable benefits in a broad array of domains, particularly for youth who are not otherwise getting their mental health needs met.

Acknowledgments
The authors would like to thank Danielle Arntson, Jeremy Astesano, Maria Bandriwsky, Turner Block, Abigail Blum, Marie Chamberlain, Loan Ho, Cherrelle Jones, Lauren Nowakowski, Jean Rhodes, Emily Romero, Geert-Jan Stams, Allison Tetzlaff, and Charlotte Utschig. This research was supported by funding from the Center for Evidence-Based Mentoring.

Authors’ Contributions
CC and ER conceptualized the study, oversaw sample selection and coding, and drafted the manuscript. KB, SB, MH, NF, and KC coded studies for analysis and contributed to the manuscript. MA ran analyses and contributed to manuscript writing.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Additional search strategy details. [PDF File (Adobe PDF File), 196 KB-Multimedia Appendix 1]

Multimedia Appendix 2
Selected characteristics of 80 studies evaluating 83 mobile technology-delivered interventions for youth. [PDF File (Adobe PDF File), 402 KB-Multimedia Appendix 2]

References


151. Hilt LM, Swords CM. Mindfulness mobile app for rumination adolescents: a randomized controlled trial. In: Proceedings of the Anxiety and Depression Association of America Annual Conference. 2022 Presented at: Anxiety and Depression Association of America annual conference; 2022; Denver, CO.


Marx LS. A mindful eating "App" for non-treatment-seeking university students with eating and weight concerns. ProQuest. 2016. URL: https://www.proquest.com/docview/1896581612/abstract/8306C60B5D4F1AD5C0?accessed=2021-10-02


Nolan J. Effectiveness, Feasibility, and Acceptability of a Mindfulness Based Mobile Application for Undergraduate Health Science Students. ProQuest. 2019. URL: https://www.proquest.com/openview/80ad66e2794d4571056ed01e89f2902/1?pq-origsite=gscholar&cbl=51922&diss=y [accessed 2021-10-02]


https://mental.jmir.org/2022/7/e34254

JMIR MENTAL HEALTH
Conley et al

Vol 9 · Iss 7 · e34254 · p. 30

(page number not for citation purposes)


Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>effect size</td>
</tr>
<tr>
<td>mTDI</td>
<td>mobile technology–delivered intervention</td>
</tr>
<tr>
<td>PRISMA</td>
<td>Preferred Reporting Items for Systematic Reviews and Meta-Analyses</td>
</tr>
<tr>
<td>TDI</td>
<td>technology-delivered intervention</td>
</tr>
<tr>
<td>VR</td>
<td>virtual reality</td>
</tr>
</tbody>
</table>

©Colleen S Conley, Elizabeth B Raposa, Kate Bartolotta, Sarah E Broner, Maya Hareli, Nicola Forbes, Kirsten M Christensen, Mark Assink. Originally published in JMIR Mental Health (https://mental.jmir.org), 29.07.2022. 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 Mental Health, is properly cited. The complete bibliographic information, a link to the original publication on https://mental.jmir.org/, as well as this copyright and license information must be included.