Tailoring acceptance and commitment therapy skill coaching in-the-moment through

smartphones: Results from a randomized controlled trial

Michael E. Levin [a*], Jack Haeger [a] & Rick A. Cruz [a]

[a] Utah State University, Department of Psychology, 2810 Old Main Hill, Logan, UT 84322.

* Corresponding author. Utah State University, 2810 Old Main Hill, Logan, UT 84322, United

States. Phone: +001 (541) 531-3892; Fax: +001 (435) 797-1448, E-mail address:

Michael.Levin@usu.edu.

Abstract

There is growing evidence for the efficacy of acceptance and commitment therapy (ACT) interventions delivered through smartphones, but research has not yet focused on how to optimize such interventions. One benefit of mobile interventions is the ability to adapt content based on in-the-moment variables. The current randomized controlled trial evaluated whether an ACT app that tailored skill coaching based on in-the-moment ecological momentary assessments (EMAs) would be more efficacious than the same app where skill coaching was random or an EMA-only condition. A sample of 69 adults interested in using a self-help app were randomized to one of three app conditions and used the app for the following four weeks. Results indicated equivalently high user satisfaction with the tailored versus random apps. Participants used the EMA-only app the most and the tailored app the least, but overall adherence was adequate. Participants in the tailored app improved significantly more on psychological distress and positive mental health relative to the random app and EMA-only conditions. However, no differences were found between the random app and EMA-only conditions on outcomes. Between group differences over time were also found on psychological inflexibility, but this appeared to be primarily due to a lower rate of improvement in the random app condition relative to both tailored and EMA-only. Overall, these results suggest that tailoring ACT skill coaching based on in-the-moment variables leads to greater efficacy.

Keywords: mHealth; psychological inflexibility; mindfulness; micro-interventions; just-in-time adaptive interventions.

Tailoring acceptance and commitment therapy skill coaching in-the-moment through

smartphones: Results from a randomized controlled trial

Smartphones are a promising platform for delivering self-guided mental health

interventions. Most adults in the United States own a smartphone (77%), including among low

income populations (64% with income < $30,000; Pew Research Center, 2017). There is a large

literature indicating the efficacy of computerized self-guided interventions more broadly (see

Andersson, 2016 for a review), with more preliminary research suggesting such interventions are

also efficacious when delivered through smartphones (Torous, Levin, Oser & Ahern, 2017).

There are unique advantages to delivering interventions through smartphones due to their

ready availability to assess and deliver mental health interventions throughout people's day.

Mobile apps can deliver low intensity, high frequency interventions (Heron & Smyth, 2010),

which may be ideal for supporting the strengthening and generalization of psychological skills

(Levin, Haeger, Pierce & Cruz, 2017a). Furthermore, the decision to provide an intervention or

not through an app and the type of intervention provided can be tailored based on in-the-moment

variables to further enhance interventions (i.e., Just-In-Time Adaptive Interventions [JITAI];

Nahum-Shani et al., 2018). Theoretically such adaptive tailoring of mobile interventions might

produce more efficient and effective interventions, but there is limited research to-date.

A substantial body of research already demonstrates the efficacy of tailoring the content

of online interventions more broadly based on variables such as baseline user characteristics,

rather than in-the-moment variables relevant to JITAIs (e.g., Lustria et al., 2013; Strecher et al.,

2008). Recently, research has begun to emerge suggesting the potential benefits from adaptive

tailoring of *when* brief app-based interventions are delivered for problems including sedentary

behavior (Bond et al., 2014) and general stress (Smyth & Heron, 2016). The most direct

evidence of adaptive tailoring for *when* interventions are delivered was found in a randomized

controlled trial (RCT) for stress reduction comparing ecological momentary assessment (EMA)

alone, EMA plus random intervention delivery (intervention occurred at random times), and

EMA plus tailored delivery (intervention occurred specifically at times of high reported stress;

Smyth & Heron, 2016). Results indicated the tailored app condition produced greater effects than

the random app condition, both of which outperformed EMA-only. However, except for a pilot

study (Levin et al., 2017a), we have not found any published studies testing the impact of

tailoring *what* intervention is delivered through a smartphone based on in-the-moment variables,

which is another critical component of adaptive tailoring.

Existing treatment models offer one method for informing adaptive tailoring of what

interventions are delivered to users. The current study evaluated a tailored mobile app based on

acceptance and commitment therapy (ACT; Hayes, Strosahl & Wilson, 2011). The ACT model

specifies a set of key treatment components including acceptance (willingness to experience

aversive internal emotions without avoidance), cognitive defusion (relating to thoughts as just

thoughts), present moment awareness (flexible attention to current experiences), and values

(clarifying personally meaningful qualities of action; Hayes et al., 2011). Theoretically, each

ACT component impacts a specific aspect of psychological inflexibility, the key target in ACT in

which behavior is rigidly governed by internal experiences, rather than values or direct

contingencies (Hayes, Pistorello & Levin, 2012). Consistent with this theory, a large body of

laboratory-based component research supports the isolated effects of ACT's acceptance,

defusion, values, and present moment awareness components (Levin, Hildebrandt, Lillis &

Hayes, 2012). Thus, the ACT model might specify a key set of variables to assess in-the-moment

to guide selection of targeted ACT component skills: using cognitive defusion skills when fused

with thoughts, values skills when disconnected from values, present moment skills when on

"autopilot" and acceptance skills when avoiding/struggling with emotions. Yet, this is an

empirical question since minimal research to-date has tested the efficacy of tailoring ACT

components to target corresponding pathological processes, and there are questions regarding the

degree to which ACT components overlap versus have distinct functions for tailored delivery.

Preliminary RCTs suggest ACT can be effectively delivered in mobile app formats

(Torous et al. 2017). However, only one published study has evaluated the impact of tailoring

ACT skills in a mobile app (Levin et al., 2017a). In an open trial, 14 depressed and anxious

clients currently receiving ACT from a therapist used the ACT Daily mobile app for two weeks.

The app prompted EMA check-ins 3 times a day, with the option to receive a tailored skill based

on one's check-in responses. The ACT Daily app provided tailored skills for four of the six key

ACT components (i.e., acceptance, defusion, present moment, values). ACT Daily excluded two

ACT components, self-as-context and committed action, due to the prototype nature of the app,

difficulty operationalizing in-the-moment assessment and tailored skills for these components,

and their more limited component evidence base (Levin et al., 2012). In this pilot evaluation of

ACT Daily, significant improvements were found over the two-week app testing period on

depression and anxiety as well as psychological inflexibility. Furthermore, participants

demonstrated significant improvements on check-in variables immediately after each skill

coaching session, with targeted effects found for the acceptance and present moment skill

components. For example, acceptance skills had a greater impact on in-the-moment levels of

"fighting feelings" relative to other ACT component skills, supporting the importance of using

acceptance skills specifically when high in non-acceptance. Overall, this pilot trial suggests an

ACT app that tailors what component skills are provided based on in-the-moment responses

could be efficacious. However, this was a small, uncontrolled trial and thus it is difficult to determine the isolated effects of the app in general, and of the tailoring feature more specifically.

The current study aimed to build on preliminary research examining the impact of tailoring what ACT skills are provided to users based on in-the-moment assessment responses. A RCT was conducted with a sample of 69 adults expressing interest in using a self-help mobile app. A broad sample was included to be consistent with ACT's transdiagnostic approach across levels and forms of distress (Hayes et al., 2012; Levin, Haeger, Pierce & Twohig, 2017b) as well as the range of individuals who might access a general self-help app. The study specifically compared three conditions using a mobile app over a four week period: a version of the ACT app in which the specific ACT component skills provided were tailored based on in-the-moment check-in responses (tailored app condition), a version of the app with the EMA check in, but in which skills were random rather than tailored (random app condition), and a version where users only completed the EMA check-in without any skill coaching (EMA-only condition). We predicted that participants using the tailored app would improve the most over the four week testing period on measures of psychological functioning and psychological inflexibility, with the random app condition having a somewhat weaker effect though still greater than EMA-only.

Method

Participants

A sample of 69 adults were recruited for the study from January 2017 to April 2017. Eligibility criteria were living in the US, being fluent in English, 18 years of age or older, owning an Android or iPhone, and interested in participating in a mobile self-help study. Recruitment was primarily conducted at a mid-sized university in the Mountain West region of the United States (e.g., flyers, online advertisements). Potential participants were directed to an

online screener for various clinical trial opportunities offered through the laboratory. Although additional recruitment efforts for the clinical trial screener were conducted nationally (through a website, professional listserv postings, community flyers), all but one participant was a student from the local university. Four additional people completed informed consent, but dropped out for unknown reasons prior to starting baseline assessment (see Figure 1 for participant flow).

Participant age ranged from 18 to 46 years with a median age of 20 ($M = 21.9$, $SD = 5.47$). The majority of participants were female (68.1%) and White (94.2%), with the remaining 2.9% Asian, 1.4% American Indian/Alaska Native, and 1.4% Black/African American. With regards to ethnicity, only 5.8% identified as Hispanic or Latino. A minority of participants (19%) were married or cohabitating, with 13% employed full time and 26% employed part time.

The study purposefully did not require particular mental health symptoms to be included in the study. This is consistent with transdiagnostic applications of ACT to a wide range of problems as well as non-clinical samples, irrespective of level or form of distress (Hayes et al., 2012; Levin et al., 2017b). The majority of the sample (62.3%) did report moderate or higher rates of depression, anxiety, or stress symptoms at baseline, according to recommended cutoff scores from the Depression, Anxiety and Stress Scale (DASS; Lovibond & Lovibond, 1995).

Procedures

A RCT design was used with participants assigned to one of three conditions for the four week intervention period (tailored app, random app, or EMA-only). No changes were made to the trial after data collection commenced. All procedures were completed online through email, text message, and the Qualtrics research platform (see Figure 1 for an overview of procedures).

Participants first completed online informed consent and then an online baseline assessment. After the baseline assessment, participants were automatically randomized to one of

three ACT Daily app conditions (tailored app, random app, or EMA-only). Randomization was

conducted through the Qualtrics randomizer feature in blocked sets of 15 to help ensure minimal

imbalance in sample sizes between conditions. Participants were partially blinded to condition

assignment as the EMA-only condition was clearly discernable to those using the app. However,

the random versus tailored app conditions were nearly identical (including check-ins and options

to access skill coaching) and participants were blinded to assignment to these two conditions.

After completing baseline and being assigned to study condition, participants assigned to

the tailored and random app conditions were provided a 15-minute online orientation on how to

use their assigned app over the next four weeks and an introduction to the ACT skills taught in

the app (participants in the EMA-only condition did not complete the orientation). They were

instructed to complete the app check-in twice a day when prompted by text message, with the

option to complete additional check-ins at any time. After four weeks, participants completed an

online post assessment. After completing the post assessment, participants in the EMA-only

condition were provided access to the tailored app. The only participation incentives were

research credits for completing each of the three assessments among students enrolled in relevant

university courses (research credit was not provided contingent on use, or not, of the mobile

app). The study received necessary ethical approval through the Utah State University

Institutional Review Board. Informed consent was obtained for all participants.

Mobile app platform. Non-native apps (only accessible online) were developed in

Qualtrics for this research study. Although Qualtrics is primarily designed for surveys, we have

found in past studies that it is also effective at delivering engaging online self-help interventions

(e.g., Levin et al., 2017b). Qualtrics provides a sophisticated array of interactive features,

tailoring based on responses, and multimedia delivery options, all within a responsive design that

supports visually appealing, easy-to-complete coaching sessions on a smartphone. The ACT Daily app included a check-in assessment, library of skill coaching sessions (see table 1 for examples), and tailoring rules for what ACT component was provided (depending on study condition). The details of the app components are described below and outlined in Figure 2.

EMA check-in assessment (provided to all three conditions). Text messages prompting participants to check-in with an app URL were sent by a research assistant semi-randomly twice a day over the four week app intervention period. Prompts were spaced at least an hour apart, with one between 9am and 3pm and the other between 3pm and 9pm each day. The timing of messages was randomized using a random clock time generator feature (random.org). Participants were also instructed on how to access the non-native app at any time.

In all three conditions, participants were first directed to a check-in assessment each time the app was accessed. The check-in included 7 items on a visual analogue scale from 0 (none) to 100 (extremely). Items included: 1) feeling sad/depressed, 2) feeling anxious/afraid, 3) unable to do what matters, 4) fighting your feelings, 5) stuck in thoughts, 6) running on autopilot, and 7) disconnected from values. Each item was explained during the online orientation to the app. A previous pilot using these items found support for their validity in assessing distinct aspects of psychological processes relevant to ACT skills (Levin et al., 2017a). The first three items (sadness, anxiety, unable to do what matters) were proximal outcomes and not used for tailoring.

EMA-only condition. Participants in the EMA-only condition simply completed the app check-in assessments when prompted twice daily by text message. No coaching skills or other app content was provided besides these check-in assessments.

Online app orientation (tailored and random app conditions). After being assigned to study condition, participants assigned to the tailored or random app conditions were asked to

complete a 15-minute online orientation through Qualtrics. The orientation was required before

receiving access to the app. The orientation included clarifying personal goals for using ACT

Daily, an introduction to each of the four ACT Daily component skills, and instructions for how

to begin using the app. To identify personal goals for using ACT Daily, participants were guided

through a brief set of interactive worksheets to reflect on personal values, important areas they

want to work on in their life, and what psychological barriers get in the way. To introduce each

ACT component, participants received two pages of didactic, text describing what the

component is, as well as completing a brief introductory exercise for each component.

        Skill coaching in tailored and random conditions. Participants in the tailored or

random app conditions were provided three options after completing the check-in assessment –

quick skill, depth skill, or end the session. Selecting "quick skills" immediately directed

participants to a randomly selected, one-page text instructed skill to practice in the moment for a

given ACT component library, with an option to see another quick skill or end the session.

Selecting "another skill" would randomly present a different quick skill from the relevant skill

library, with the option to select another skill or end the session. Selecting "depth skills" directed

participants to a menu for the given ACT component library with a list of longer ACT skill

training sessions to choose from for that treatment component, including multipage interactive

exercises and audio-guided exercises. Selecting "end session" simply closed the session without

any skill training (for users who did not want to practice a skill in the moment).

        Quick and depth skills were organized by treatment component: acceptance, defusion,

present moment, and values. Users were given access to skills from only one ACT component

for each app session, with 28 quick skills and 6 depth skills available per component. A total of

136 skill sessions were provided to ensure users continued to find a variety of new skills to stay

engaged over time (users found the previous prototype with 28 skills to be very redundant after 2 weeks; Levin et al., 2017a). Examples of skill coaching sessions are provided in Table 1. After completing a skill coaching session, users were prompted to re-complete the check-in assessment to assess changes in variables. Similar to the previous ACT Daily prototype (Levin et al., 2017a), this app did not include the ACT self-as-context and committed action components, which require further research in operationalizing in-the-moment assessment and tailored interventions.

Tailored skill coaching condition. In the tailored condition, participants were assigned an ACT component skill set based on the check-in answers that were just provided. Participants received the ACT component linked to the highest-rated problem in the moment. Participants were directed to acceptance skills if they rated "fighting your feelings" the highest, defusion if "stuck in thoughts" rated highest, present moment if "running on autopilot" rated highest, and values if "disconnected from values" rated highest. Thus, skills were tailored to address their most relevant challenge based on the psychological inflexibility model. The anxiety, sadness, and unable to do what matters check-in items were not factored into tailored decision making.

Random skill coaching condition. In the random condition participants completed the EMA check-in, but then received a randomly selected ACT skill component when they selected "quick skill" or "depth skill." The skill coaching they received was not associated in any way with check-in responses. The tailored and random skill coaching conditions were identical in all other ways besides whether the component skill set provided was based on check-in responses.

Researcher contact. This research study was mostly automated, with researcher contact limited to email and text messages to troubleshoot technical issues and prompt completion of study procedures. These primarily consisted of standardized, general messages including text

messages to complete an app check-in and follow up email prompts if a participant stopped using the app or did not complete a study procedure (e.g., post assessment).

Measures

Depression, Anxiety and Stress Scale (DASS; Lovibond & Lovibond, 1995). The 21-item DASS assesses symptoms of depression, anxiety, and stress and was used as the primary outcome measure. All items were summed for a total score, which is consistent with research indicating subscales load onto a higher order factor of general distress (Henry & Crawford, 2005). The DASS has displayed good reliability and validity in previous research (Lovibond & Lovibond, 1995) and is sensitive to detecting ACT app intervention effects (Levin et al., 2017a). The DASS total score had excellent internal consistency in the current study ($\alpha = .92$).

Mental Health Continuum – Short Form (MHC-SF; Keyes, 2005). The 14-item MHC-SF was used as a secondary measure of positive mental health. The MHC-SF assesses aspects of wellbeing including positive affect, satisfaction with life, social integration, personal growth, autonomy, and purpose in life. The MHC-SF has been found to be reliable and valid (Keyes, 2005) and sensitive to detecting the effects of online ACT interventions (e.g., Levin et al., 2017b). Within the current study, the MHC-SF had excellent internal consistency ($\alpha = .94$).

Satisfaction with Participation in Social Roles (SPSR; Hahn et al., 2010). An 11-item set of satisfaction in social roles items from the patient-reported outcomes measurement information system (PROMIS) was used as a secondary outcome to assess social functioning. Previous research indicates that the SPSR is reliable and valid (Hahn et al., 2010) and sensitive to treatment effects (Hahn et al., 2016), but has not been evaluated for sensitivity to detecting ACT effects. The SPSR had excellent internal consistency in the current study ($\alpha = .92$)

Acceptance and Action Questionnaire-II (AAQ-II; Bond et al., 2011). The 7-item AAQ-II assessed psychological inflexibility, the primary process of change in ACT. Previous studies have found adequate reliability and validity for the AAQ-II (Bond et al., 2011). However, online ACT research has had mixed results with the AAQ-II detecting treatment effects (Levin et al., 2017b). The AAQ-II had excellent internal consistency in the current study (α = .94).

Comprehensive Assessment of Acceptance and Commitment Therapy (CompACT; Francis, Dawson & Golijani-Moghaddam, 2016). The 23-item CompACT was used as an additional measure of psychological inflexibility. The CompACT is a recently developed measure designed to assess the full range of inflexibility processes including openness to experience (acceptance/defusion), awareness, and valued action. Preliminary research indicates the CompACT is reliable and valid (Francis et al., 2016), although there is no published research on its sensitivity to detecting ACT effects. To minimize number of analyses, only the CompACT total score was used in the current study, which had adequate internal consistency (α = .85).

System Usability Scale (SUS; Brooke, 1996). The 10-item SUS assessed perceived program usability and satisfaction among participants assigned to the tailored or random skill coaching app conditions. The SUS has been used extensively in online research (Bangor, Kortum & Miller, 2008). The SUS had acceptable internal consistency in the current study (α = .78).

Data analyses

The planned sample size ranged from 60 to 90 participants depending on enrollment rates in relation to a planned deadline of May 2017, which would afford between 20 to 30 participants per condition. The trial was closed due to recruiting a sufficient sample (n = 69) by the deadline, which provided adequate power (.80) to detect a significant ($p < .05$) time by condition interaction for three study conditions with an effect size of Cohen's d = .38.

To examine program acceptability, descriptive statistics for program satisfaction and engagement were calculated in the two skill coaching app conditions. Independent sample t-tests tested whether program satisfaction rates differed between the tailored and random app conditions and whether program usage rates differed between all three conditions. Chi square tests were conducted to compare rates of post assessment completion between conditions.

Prior to examining treatment effects, analyses tested for any between condition differences on process or outcome measures between conditions at baseline using a series of one-way ANOVAs. Given the sample size, post hoc analyses were conducted using Tukey least significant difference tests, whether or not omnibus ANOVA effects were significant, to further test baseline differences. A series of binary logistic regressions and ANOVAs examined whether post data was missing at random or associated with baseline variables (demographics, outcomes, processes of change, or study condition). Skewness and kurtosis were examined for each variable to assess whether variables were normally distributed. Q-Q plots tested whether residuals for each mixed-model repeated-measures (MMRM)) analysis were normally distributed.

MMRM with unstructured covariance matrices were used to test for time by condition effects on each outcome and process measure using an intent-to-treat (ITT) approach. MMRM models missing data, which allowed for inclusion of the full sample randomized to treatment condition, irrespective of whether post assessment was completed. Any baseline differences between conditions were included as covariates. Planned post hoc analyses tested differences between conditions at post and within condition improvements from pre to post. Cohen's d effect sizes were calculated for time by condition effects (using the F statistic and df) and post hoc comparisons (using mean difference, variance, and covariance estimates) based on recommended procedures (Verbeke & Molenberghs, 2000; Wackerly, Mendenhall, & Scheaffer, 2008).

Results

Program Satisfaction

Program usability ratings were examined among participants in the random and tailored app conditions who completed the post assessment ($n = 37$, 82% of the sample), of which, 92% ($n = 34$) completed at least one skill coaching session in the app. Usability ratings were high and equivalent between conditions (Tailored SUS $M = 83.13$, $SD = 11.19$; Random SUS $M = 87.88$, $SD = 10.19$; $t(31) = 1.28$, $p = .21$). These SUS scores were consistent with a "good" to "excellent" rating based on scale benchmarks (Bangor et al., 2008) and were similar to the high usability scores found in the first evaluation of ACT Daily (SUS $M = 89.08$; Levin et al., 2017a).

Participants in both skill coaching conditions also equally agreed on a 6-point scale (4 indicating "slightly agree") that "The ACT Daily check-ins seemed to guide the skill coaching I received" (Tailored app $M = 4.44$, $SD = 1.15$ vs. Random app $M = 4.53$, $SD = 1.13$), that "It was important that the ACT Daily check-in guides the specific skill coaching I receive" (Tailored app $M = 4.94$, $SD = 1.00$ vs. Random app $M = 4.88$, $SD = .93$), and that "ACT Daily was helpful to me" (Tailored app $M = 4.31$, $SD = 1.49$ vs. Random app $M = 4.59$, $SD = 1.28$). Independent sample t-tests indicated no differences between conditions on these satisfaction item ratings ($p > .10$). Thus, participants in both app conditions reported similar high satisfaction ratings and a perception that the ACT Daily app was tailoring skill coaching based on check-in responses (despite such tailoring not actually being provided in the random app condition).

Program Engagement

Over the 28-day period, participants in the EMA-only condition completed an average of 46.71 check-in assessments ($SD = 13.86$, 83% of the 56 prompted assessments). This was significantly higher than the number of check-ins completed in the tailored app condition ($M =$

25.39, SD = 20.17, 45% of prompts; $t(45)$ = 4.24, $p < .001$, d = 1.23) and the random app

condition (M = 35.86, SD = 18.99, 64% of prompts; $t(44)$ = 2.23, p = .03, d = .65). Participants in

the random app condition completed more skill training sessions (M = 25.36, SD = 17.94)

relative to tailored app (M = 13.57, SD 11.57), $t(43)$ = 2.63, p = .01, d = .78. Given the

availability of 136 skills in the ACT Daily app, this suggests participants were unlikely to

regularly repeat skill sessions. Participants in the tailored condition were significantly more

likely to not complete any skill training sessions (22%) than the random condition (0%), χ^2=

6.62, p = .02. After two weeks (halfway through the intervention), 17 participants were still using

the tailored app (74%), 18 using the random app (82%), and 23 using the EMA-only app (96%).

After three weeks, 16 were using the tailored app (70%), 17 were using the random app (77%),

and 22 were using the EMA-only app (92%). Overall app usage was lowest in the tailored and

highest in the EMA-only condition.

Preliminary analyses for MMRM

All variables were normally distributed based on skewness and kurtosis. Q-Q plots

indicated that residuals were normally distributed for each MMRM model. In terms of missing

data, 86% of participants completed the post assessment, with no significant differences in

dropout rates between conditions (see Figure 1). Missing data at post was not significantly

predicted by any baseline demographic, outcome, or process measures ($p > .10$).

A series of one-way ANOVAs indicated no significant differences between conditions on

outcome or process measures at baseline ($p > .10$). However, a post hoc analysis did indicate that

participants in the tailored app condition had significantly lower positive mental health relative

to the EMA-only condition (M_{Diff} = 7.70, SE = 3.71, p = .04). Thus, MMRM results included

baseline positive mental health as a covariate. Including positive mental health as a baseline

covariate led to minimal differences in results relative to MMRM analyses without this covariate

(Cohen's *d* effect sizes for time by condition and post hoc tests differed by .05 at the most).

Between group effects on treatment outcomes

MMRM tested time by condition interactions on each outcome with the full ITT sample,

controlling for baseline positive mental health. Estimated marginal means (modeling missing

data and covariate effects) for each condition and time point are provided in table 2 and MMRM

results are provided in table 3.

Significant time by condition interactions were found for DASS psychological distress

(see Figure 3) and MHC-SF positive mental health with a statistical trend for SPSR social

functioning ($p = .08$), with small to medium effect sizes ranging between .42 and .56. Post hoc

analyses indicated that in each case only participants in the tailored app condition significantly

improved over time on outcomes, with medium effect sizes ranging between .53 and .73. In

contrast, participants in the random app and EMA-only conditions did not significantly improve

over time. The effect sizes for the random and EMA-only conditions were negative, however all

confidence intervals included zero (see Table 2). At post, participants in the tailored app

condition had significantly lower distress and higher social functioning relative to the random

app and EMA-only conditions, with large effect sizes ranging between .85 and 1.25. There were

no statistically significant differences between the random app and EMA-only conditions at post.

Between group effects on treatment processes of change

MMRM results also tested for time by condition effects on psychological inflexibility

measures, controlling for baseline positive mental health (see Tables 2 and 3). A significant time

by condition effect was found for psychological inflexibility as measured by the CompACT ($d =$

.54) with a statistical trend also found for the AAQ-II ($p = .06$, $d = .45$).

Post hoc analyses on the CompACT indicated that participants in both the tailored app condition and EMA-only condition significantly improved over time with a medium effect size (d = .63 and .57 respectively), while there was no change over time in the random app condition. Participants in the random app condition were significantly more psychologically inflexible (CompACT) at post relative to the tailored app (d = 1.23) and EMA-only conditions (d = 1.02). There was no difference between the tailored and EMA conditions on CompACT scores at post.

Post hoc analyses on the AAQ-II indicated that only participants in the tailored app condition significantly improved over time on psychological inflexibility (d = .62), with effect sizes near zero for the random app (d = -.10) and EMA-only conditions (d = -.03). There were no differences between conditions at post on AAQ-II scores besides a trending medium effect size for lower psychological inflexibility in the tailored relative to random app condition (d = .76).

Discussion

This study evaluated whether a mobile app that tailored ACT skills based on in-the-moment assessment responses would lead to greater effects on mental health and psychological inflexibility. Consistent with predictions, participants receiving tailored ACT skill coaching improved more on distress, positive mental health, and social functioning over time relative to an EMA-only condition and a condition where skill coaching was not tailored. Surprisingly, despite equivalent satisfaction ratings and higher engagement rates than the tailored app, the random app performed equivalently to EMA-only and did not demonstrate any within condition improvements over time. Results were somewhat mixed for processes of change, with between condition differences on psychological inflexibility due primarily to a lack of improvement on inflexibility in the random app condition. Overall, results indicate that tailoring what ACT skills are provided based on in-the-moment assessment data may improve mental health outcomes.

This is the first study of which we are aware that specifically tested tailoring *what* skills are provided in a mobile app based on in-the-moment data. This study adds to the broader literature on just-in-time adaptive interventions (JITAI; Nahum-Shani et al., 2018), which aims to adaptively tailor *when* and *what* interventions are provided based on in-the-moment data. JITAI research thus far has focused on studying tailoring *when* interventions are delivered, with promising preliminary findings (e.g., Bond et al., 2014; Smyth & Heron, 2016). This study suggests that tailoring *what* intervention is delivered is also important.

A preliminary tailoring approach was used in which skills were categorized into four isolated ACT components (acceptance, defusion, present moment, values) and selected based on whichever associated inflexibility variable was rated highest. This assumes each ACT component functions differentially and is more effective in targeting their associated inflexibility process. The positive findings with the tailored app support this approach. However, it is unclear whether more sophisticated tailoring might improve outcomes, such as ideographically defined tailoring rules (for subgroups of users) or tailoring based on additional assessment variables (including passive data such as GPS). Given the overlap in ACT components, it is also unclear whether tailored content should focus on delivering isolated ACT components (e.g., defusion alone) versus a mix of ACT components tailored to a current challenge. Furthermore, ACT Daily prompted participants semi-randomly, rather than adaptively prompting *when* to use ACT skills. Outcomes might be improved with a full JITAI approach in which both *when* and *what* skills were tailored. Thus, this study provides "proof of concept" for the benefits of tailoring an ACT app, pointing to an array of future research questions to optimize such an approach.

The random coaching condition helps rule out a number of alternative explanations for findings given the app was identical in all ways besides whether skill tailoring was provided.

This is further supported by findings that relative to the tailored condition, participants in the random condition were equally satisfied with the random app, equally believed the app tailored skills, and actually used the random app more. This reduces the potential for between group findings to be due to extraneous variables such as demand characteristics, placebo, or dosage.

It was surprising how inert the random condition was, failing to outperform the EMA-only condition and potentially even performing worse than EMA-only on psychological inflexibility. This suggests that randomly providing ACT skills irrespective of what components are most relevant may be ineffective and some degree of tailoring to the moment is needed (or at least an option for users to choose their own skills). This is consistent with ACT's ideographic, present-focused approach (Hayes et al., 2011) and provides empirical support for the importance of matching interventions to current context. The inert effects for the random app may have limited generalizability to other self-guided ACT apps given their broad empirical support in previous studies (Torous et al., 2017). There may be unique features to the ACT Daily app or sample that led to inert effects from the random version of the app (e.g., brevity of ACT skill training, lack of choice with ACT component skills, use of an unassociated EMA check-in).

Another surprising finding was that although the tailored app condition produced greater improvements than the random app, participants used the tailored version less often. This differs from previous studies demonstrating greater usage of ACT apps relates to greater improvements in psychological inflexibility (e.g., Mattila et al., 2016). One possibility is that tailoring ACT skills produced improvements more efficiently by matching skill coaching to in-the-moment needs (so that users benefit more from fewer sessions), while participants in the random condition continued to access skills because of a lack of improvements. In retrospect, we would have predicted higher usage of the tailored app, due to greater perceived helpfulness. Future

research is needed to examine whether tailoring improves the efficiency of ACT skill coaching, and whether increasing frequency of use of tailored skill coaching improves outcomes.

This study has the most direct implications for mobile apps, indicating the potential importance of tailoring skill coaching based on in-the-moment variables. Not only does this highlight an area for future app development, but it also may inform an evidence-based feature that clinicians might look for when selecting an app to use with clients (Torous et al., 2017).

A broader and more tenuous generalization is that these findings might also have some relevance to face-to-face therapy. Currently, there is a lack of empirical guidance for more fine-grained clinical decision making with therapies such as ACT, such as when to provide what treatment components to a client (Levin, Herbert & Forman, 2017). There are substantial challenges to even generating a basic amount of research evidence when various factors are considered (what client, under what conditions, in relation to what processes, etc…). Research on mobile apps will always have a notable gap with face-to-face clinical decision making, but might provide initial guides for addressing such clinical questions. These areas might also be bridged through adjunctive mobile technologies that could eventually assist clinicians in such evidence-based decision making on what skills to target in a session.

Limitations and Future Research

There were limitations that affect the generalizations that can be made from this study. The sample had notable heterogeneity with regards to clinical concerns (or lack thereof) and homogeneity with regards to a primarily white, college sample. Results may not generalize to the potential differences and needs of clients seeking clinical services or with specific diagnoses.

Additional limitations may affect reproducibility of results. The study was small with 69 participants, which increased the potential for Type I and Type II error. Although differences

were found between the random and tailored app, the study was underpowered to detect

differences between active conditions. The clinical trial was not registered, which further limits

certainty with regards to reporting all relevant data under pre-specified parameters and analyses.

A final arguable limitation is the use of Qualtrics to deliver a non-native app. The use of

Qualtrics means that the app cannot be directly offered on the app store, which eliminates an

appealing aspect of mHealth research in which efficacious interventions can be readily

disseminated in the exact same format after testing. However, Qualtrics allowed for the rapid

development and testing of multiple app conditions with advanced interactive features, without

grant funding. Thus, Qualtrics provides a way to efficiently conduct innovative research, without

significant delays in an increasingly funding-limited environment. Results from this study

suggest that Qualtrics can be used effectively to conduct such experimental research.

Overall, this study highlights a promising way to combine innovative features of mobile

apps with theory-based research on clinical models to both deliver more effective computerized

interventions and increase our understanding of underlying treatment models. The current study

found that tailoring what ACT component is provided in-the-moment makes a difference in

treatment outcomes. This bolsters the current evidence for the utility of these ACT components

and clarifies a pathway for component research to directly inform future app interventions. With

continued efforts, such research may ultimately help inform both more effective mobile apps as

well as face-to-face clinical decision making, helping technologies and therapists know when to

provide what treatment strategy with a client based on in-the-moment variables.

Compliance with Ethical Standards

Conflict of interest statement: Dr. Levin is a research associate with Contextual Change LLC, a small business that focuses on developing commercial online programs to address college student mental health. Mr. Jack Haeger and Dr. Rick Cruz declare that they have no conflicts of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved by the Utah State University Institutional Review Board.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Author contributions. MEL: Was the principal investigator including designing and overseeing conduct of the study, conducting data analyses, and writing the manuscript. JH: Assisted with design of the study, was the primary research coordinator for day-to-day operation of the study, and assisted in preparing the manuscript. RAC: Was a co-investigator who assisted in designing the study, conducting data analyses, and preparing the manuscript.

<div align="center">**References**</div>

Andersson, G. (2016). Internet-delivered psychological treatments. *Annual Review of Clinical Psychology, 12,* 157-179.

Bangor, A, Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction, 24*, 574-594.

Bond, D.S., Thomas, J.G., Raynor, H.A., Moon, J., Sieling, J., Trautvetter, J., … & Wing, R.R. (2014). B-MOBILE – a smartphone-based intervention to reduce sedentary time in overweight/obese individuals: A within-subjects experimental trial. *PLoS ONE, 9*, e100821.

Bond, F. W., Hayes, S.C., Baer, R. A., Carpenter, K., Orcutt, H. K., Waltz, T., & Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: A revised measure of psychological flexibility and acceptance. *Behavior Therapy, 42*, 676-688.

Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In P.W. Jordan, B. Thomas, B.A. Weerdmeester, and I.L. McClelland (Eds.) *Usability Evaluation in Industry* (189-194). London: Taylor and Francis.

Francis, A.W., Dawson, D.L. & Golijani-Moghaddam, N. (2016). The development and validation of the comprehensive assessment of acceptance and commitment therapy processes (CompACT). *Journal of Contextual Behavioral Science, 5*, 134-145.

Hahn, E.A., Beaumont, J.L., Pilkonis, P.A., Garcia, S.F., Magasi, S., DeWalt, D.A. & Cella, D. (2016). The PROMIS satisfaction with social participation measures demonstrate responsiveness in diverse clinical populations. *Journal of Clinical Epidemiology, 73*, 135-141.

Hayes, S. C., Pistorello, J. & Levin, M.E. (2012). Acceptance and Commitment Therapy as a

unified model of behavior change. *The Counseling Psychologist, 40*, 976-1002.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2011). *Acceptance and Commitment Therapy:*

The process and practice of mindful change. New York: The Guilford Press.

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress

Scales (DASS-21): construct validity and normative data in a large non-clinical sample.

British Journal of Clinical Psychology, 44, 227-239.

Heron, K. E., & Smyth, J. M. (2010). Ecological Momentary Interventions: Incorporating mobile

technology into psychosocial and health behavior treatments. *British Journal of Health*

Psychology, 15, 1-39.

Keyes, C.L.M. (2005). Mental illness and/or mental health? Investigating axioms of the complete

state model of health. *Journal of Consulting and Clinical Psychology 73*, 539–548.

Levin, M.E., Haeger, J., Pierce, B. & Cruz, R. (2017a). Evaluating an adjunctive mobile app to

enhance psychological flexibility in acceptance and commitment therapy. *Behavior*

Modification, 41, 846-867.

Levin, M.E., Haeger, J., Pierce, B. & Twohig, M.P. (2017b). Web-based acceptance and

commitment therapy for mental health problems in college students: A randomized

controlled trial. *Behavior Modification, 41,* 141-162.

. Levin, M.E., Herbert, J.D. & Forman, E.M. (2017). Acceptance and Commitment Therapy: A

critical review to guide clinical decision making. In D. McKay, J. Abramowitz & E.

Storch (Eds.) *Treatments for Psychological Problems and Syndromes* (pp. 413-432).

Wiley-Blackwell.

Levin, M. E., Hildebrandt, M. J., Lillis, J., & Hayes, S. C. (2012). The impact of treatment components suggested by the psychological flexibility model: A meta-analysis of laboratory-based component studies. *Behavior Therapy*, *43*, 741-756.

Lovibond, S. H. & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales, (2nd ed.)*. Sydney, AU: Psychology Foundation of Australia.

Lustria, M. L. A., Noar, S. M., Cortese, J., Van Stee, S. K., Glueckauf, R. L., & Lee, J. (2013). A meta-analysis of web-delivered tailored health behavior change interventions. *Journal of Health Communication, 18*, 1039–1069.

Mattila, E., Lappalainen, R., Välkkynen, P., Sairanen, E., Lappalainen, P., Karhunen, L., ... Ermes, M. (2016). Usage and dose response of a mobile acceptance and commitment therapy app: Secondary analysis of the intervention arm of a randomized controlled trial. *Journal of Medical Internet Research mHealth and uHealth, 4*, e90.

Nahum-Shani, I., Smith, S.N., Spring, B.J., Collins, L.M., Witkiewitz, K., Tewari, A. & Murphy, S.A. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health support. *Annals of Behavioral Medicine, 52*, 446-462.

Pew Research Center (2017). Mobile Fact Sheet. Retrieved June 26, 2017, from http://www.pewinternet.org/fact-sheet/mobile/

Smyth, J.M. & Heron, K.E. (2016). Is providing mobile interventions "just-in-time" helpful? An experimental proof of concept study of just-in-time intervention for stress management. *Proceedings of the IEEE Wireless Health Conference*, 89-95.

Strecher, V.J., McClure, J.B., Alexander, G.L., Chakraborty, B., Nair, V.N., Konkel, J.M., ...

      Pomerleau, O.F. (2008). Web-based smoking-cessation programs: Results of a randomized

      trial. *American Journal of Preventive Medicine, 34*, 373-381.

Torous, J. B., Levin, M. E., Ahern, D., & Oser, M. (2017). Cognitive behavioral mobile

      applications: Research literature, marketplace data, and evaluation guidelines. *Cognitive*

      and Behavioral Practice, 24, 215-225.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York:

      Springer-Verlag.

Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2008). *Mathematical statistics with*

      applications. Belmont, CA: Thomson Brooks/Cole.

Table 1. List of example ACT Daily depth and quick skills by ACT treatment component.

Component	Depth Skill Examples	Quick Skill Examples
Acceptance	Sitting with emotions: Audio guided exercise (or text-based) that walks users through how to practice acceptance with a difficult emotion.	Unwanted party guest: Metaphor on costs of focusing on getting rid of an unwelcome guest and how to enjoy the party instead, even when the guest comes up.
	Did it work?: Interactive exercise that guides users to reflect on various consequences of an away move.	Self-compassion: Prompts to relate to an emotion in compassionate ways (e.g., like a scared puppy, delicate flower).
		And vs. but: Explores changing a "but statement" to an "and statement" (e.g., "I'll call my friend AND I'm sad").
Cognitive Defusion	Leaves on a stream: Audio guided exercise (or text-based) imagining placing thoughts on leaves floating by to notice them from a defused stance.	Has this worked for me: Reflect on whether being fused with the thought has worked in the past.
	Practicing flexibility with a thought: Series of interactive exercises practicing defusion with a thought (e.g., categorizing it, labeling it, how old is it, noticing it as symbols).	Thank your mind: Instructions on how to just thank your mind for doing its job, without getting fused with it. Breaking patterns: Practice doing something even while you tell yourself you cannot do it.
Present Moment Awareness	Breathing mindfulness: Audio guided exercise (or text-based) involving mindfulness of the breath.	Observing closely: Prompts to pick an object and observe its details with curiosity.
	Tracking your attention: Interactive exercise using one's finger to track attention as it wanders from past, present, and future.	Notice 5 things: Prompts to notice 5 sights, 5 sounds, and 5 sensations. Math problems vs. sunsets: Instructions on noticing experiencing just for what they are (like a sunset) versus something to evaluate and solve (like math problems).
Values	Sweet spot: Audio guided exercise (or text-based) involving reflecting on a sweet, meaningful moment.	What's on your tombstone?: Brief exercise exploring what you would want written on your tombstone (in terms of your actions).
	Values card sort: Interactive exercise involving sorting various values based on their level of importance to the user.	One tiny step: Identifying a small valued action you could take right now. Values are like a compass: Metaphor exploring qualities of values and considering what valued direction one wants to head.

Table 2. MMRM estimated marginal means with full ITT sample.

Measure	Tailored App		Random App		EMA-Only	
	Pre *M* (*SE*)	Post *M* (*SE*)	Pre *M* (*SE*)	Post *M* (*SE*)	Pre *M* (*SE*)	Post *M* (*SE*)
Psychological Distress (DASS)	30.76 (3.98)	22.10 (4.58)	33.42 (4.35)	36.06 (4.89)	38.70 (4.13)	42.04 (4.46)
Positive Mental Health (MHC-SF)	55.30 (2.65)	61.59 (3.44)	60.32 (2.71)	60.02 (3.46)	63.00 (2.60)	61.73 (3.28)
Social Functioning (SPSR)	40.81 (1.48)	44.68 (1.81)	37.36 (1.62)	36.47 (1.91)	38.29 (1.54)	37.60 (1.72)
Psychological Inflexibility (AAQ-II)	22.92 (1.29)	19.07 (1.48)	22.43 (1.40)	23.06 (1.57)	21.37 (1.33)	21.54 (1.41)
Psychological Inflexibility (CompACT)	99.76 (3.37)	108.76 (3.64)	95.82 (3.67)	93.08 (3.91)	97.81 (3.49)	106.03 (3.52)

Note: Estimated marginal means and standard errors are based on MMRM analyses using MHC-SF as a covariate (except for the MHC-SF row). DASS and AAQ-II are scored such that higher scores indicate worse distress/inflexibility. MHC-SF, SPSR, and CompACT are scored such that higher scores indicate greater positive functioning/lower inflexibility.

Table 3. MMRM results with the full ITT sample.

Measure	Time x Condition		Pre-Post Within Condition d [95% CI]			Between Condition Post d [95% CI]		
	F	d	Tailored	Random	EMA	Tailored vs. Random	Tailored vs. EMA	Random vs. EMA
Psychological Distress (DASS)	3.48*	.50	.55* [.08, 1.02]	-.17 [-.62, .29]	-.21 [-.64, .21]	.85* [.05, 1.65]	1.21* [.33, 2.08]	.36 [-.53, 1.26]
Positive Mental Health (MHC-SF)	4.32*	.56	.73** [.26, 1.20]	-.03 [-.49, .42]	-.15 [-.57, .28]	.10 [-.52, .71]	-.01 [-.61, .59]	-.11 [-.71, .49]
Social Functioning (SPSR)	2.58†	.42	.53* [.06, .99]	-.12 [-.58, .33]	-.09 [-.52, .33]	1.25** [.46, 2.04]	1.08* [.24, 1.92]	-.17 [-1.03, .69]
Psychological Inflexibility (AAQ-II)	3.00†	.45	.62* [.15, 1.08]	-.10 [-.55, .35]	-.03 [-.45, .40]	.76† [-.05, 1.56]	.47 [-.39, 1.33]	-.29 [-1.17, .59]
Psychological Inflexibility (CompACT)	4.14*	.54	.63** [.16, 1.09]	-.19 [-.64, .26]	.57** [.15, 1.00]	1.23** [.41, 2.05]	.21 [-.68, 1.11]	-1.02* [-1.93, -.10]

Notes: †*p* < .10; **p* < .05; ***p* < .01; ****p* < .001. Time by condition tests were conducted with baseline MHC-SF as a covariate except for the analysis on positive mental health. Negative effect size scores indicate effects opposite to predictions (i.e., worsening of outcomes within conditions, tailored app post scores < random app post scores < EMA-only post scores).

Figure caption

Figure 1. Participant flow diagram.

*Figure 2. ACT Daily session flowchart. Note: *Selected ACT component for the given skill coaching session is based on check-in responses for tailored app condition and based on chance for random app condition.*

Figure 3. Changes in MMRM estimated marginal means for psychological distress (DASS) by study condition.

Assessed for eligibility and
completed informed consent
(n= 73)

Declined participation - did not start
baseline assessment (n= 4)

Completed baseline
assessment and
randomized (n= 69)

Tailored app condition (n= 23)
♦ Completed at least one app
  session (n= 18, 78%)
♦ Used the app for 3 weeks or
  more (n = 16, 70%)

Random app condition (n= 22)
♦ Completed at least one app
  session (n= 22, 100%)
♦ Used the app for 3 weeks or
  more (n = 17, 77%)

EMA-only condition (n= 24)
♦ Completed at least one app
  session (n = 24, 100%)
♦ Used the app for 3 weeks or
  more (n = 22, 92%)

Completed post assessment
(n = 18, 78%)

Completed post assessment
(n = 19, 86%)

Completed post assessment
(n = 22, 92%)

*Reasons for drop out were not provided by participants with the exception of one participant in the random app and one participant in the EMA-only app condition who stated they no longer wanted to receive the text reminders to use the app.

Texted prompt
with Qualtrics link

↓

Check-in
assessment

Select "End
Session" and
exit app

*Only option
for EMA
condition
(quick and
depth skills
unavailable)*

Select "Quick
Skill"

↓

Present 1 of 28
quick skills, with
an option to keep
trying a new skill,
for a selected ACT
component *

Select "Depth
Skill"

↓

Present a menu of
6 depth skills to
choose from for a
selected ACT
component *

Post Skill Check-In
Assessment

← Exit app