Intermittent mindfulness practice can be beneficial, and daily practice can be harmful. An in depth, mixed methods study of the “Calm” app's (mostly positive) effects

Joseph Clarke⁎, Steve Draper¹

University of Glasgow, United Kingdom of Great Britain and Northern Ireland

ABSTRACT

Objectives: Despite a weak evidence base, daily use of mindfulness-based self-help smartphone applications (apps) is said to promote wellbeing. However, many do not use these apps in the way that app developers and mindfulness proponents recommend. We sought to determine whether the “Calm” app works, and whether it does so even when it is used intermittently.

Methods: Employing a mixed-methods design, we recruited a self-selected sample of 269 students from a Scottish university (81% female, 84% white, mean age 23.89) to engage with a seven-day introductory mindfulness course, delivered using Calm, currently one of the most popular, yet under-researched, apps.

Results: Daily course engagement was associated with significant gains in wellbeing (p ≤ 0.001, d = 0.42), trait mindfulness (p ≤ 0.01, d = 0.50) and self-efficacy (p ≤ 0.014, d = 0.21). Intermittent course engagement was also associated with significant gains in wellbeing (p ≤ 0.028, d = 0.34), trait mindfulness (p ≤ 0.010, d = 0.47) and self-efficacy (p ≤ 0.028, d = 0.32). This study is therefore the first to demonstrate that the Calm app is associated with positive mental health outcomes. It also shows that regular use is not essential. A thematic analysis of qualitative data supported these quantitative findings. However it also revealed that some participants had negative experiences with the app.

Conclusions for practice: Mindfulness-based self-help apps such as Calm have the potential to both enhance and diminish users’ wellbeing. Intermittent mindfulness practice can lead to tangible benefits. Therefore, mindfulness proponents should not recommend daily practice, should increase awareness of the potential for negative outcomes, and resist the idea that mindfulness practice works for everyone. Developers of mindfulness apps ought to make specific features customisable in order to enhance their effectiveness.

1. Introduction

Mindfulness aims to cultivate a mental state of paying attention to one's sensory, cognitive, and emotional experiences “in a particular way; on purpose, in the present moment, and non-judgmentally” (Kabat-Zinn, 1994, p.4). This mindful state contrasts with the mindless state of “running on automatic-pilot” and allowing one’s mind to wander (Creswell, 2017); an unintentional yet common cognitive experience (Killingsworth and Gilbert, 2010) which often leads to rumination, anxiety, worry, fear and anger (Keng et al., 2011). There is a tendency in the literature to conflate various definitions of “mindfulness” (Chambers et al., 2009). In addition to being described as both a formal practice and a transient state, mindfulness is conceived as a way of being; a multifaceted (Lilja et al., 2013) psychological trait or disposition, which varies between individuals and refers to the tendency to be mindful in everyday life (Brown and Ryan, 2003; Bajaj et al., 2016).

Mindfulness practice is an increasingly popular (Brown et al., 2015) method of enhancing wellbeing. A recent meta-analysis of 88 randomised controlled trials (RCTs) demonstrated that regular mindfulness practice produces significant increases in self-reported trait mindfulness (Quaglia et al., 2016). High trait mindfulness is positively correlated with wellbeing (Brown and Ryan, 2003), and negatively correlated with psychological symptoms, thought suppression, difficulties with emotional regulation (Baer et al., 2006), and negative thought frequency (Freven et al., 2008).

Mindfulness is rooted in 2500-year-old Buddhist tradition (Gunaratana, 2002). The Buddhist concept of “dukkha” holds that life
inevitably involves suffering, or dissatisfaction (Teasdale and Chaskalson, 2011). To avoid dissatisfaction, we devote our energy to attaining and repeating pleasurable experiences. The problem is that all pleasurable experiences are impermanent, so we live in a repeated cycle of pleasure-seeking, which is ultimately unfulfilling (Goldstein, 2002). According to Buddhism, the “goals of mindfulness” are “to experience enlightenment, to perceive the true nature of reality” (Dimidjian and Linehan, 2003), to attain “compassion” and “wisdom” (Lykins and Baer, 2009) and “insights into important fundamental truths... which will finally lead to the ultimate goal of liberation” (Schmidt, 2011). Therefore, many proponents of mindfulness practice argue that it is “not to be understood as a simple relaxation technique” (Glück and Maercker, 2011) but rather “a regular lifelong practice in which [trait] mindfulness is cultivated over a period of many years” (Lykins and Baer, 2009).

Yet since the 1970s a secular conception of mindfulness, devoid of Buddhist morality and spirituality (Grossman, 2011), has been used to address mental health in the West. This began with the development of short-term psychotherapeutic treatments including Mindfulness-Based Stress Reduction (MBSR; Kabat-Zinn, 1982) and Mindfulness-Based Cognitive Therapy (MBCT; Segal et al., 2002). By training the mind to impartially observe and accept experience, rather than attempting to suppress or change it (as in traditional cognitive behavioural therapy) these so-called “third wave” (Hayes and Holmang, 2017) mindfulness-based therapies are thought to improve emotional regulation (Chambers et al., 2009) and enable a healthier relationship with one’s thoughts (Teasdale et al., 2000).

Recent meta-analyses demonstrate the efficacy of MBCT and MBSR in reducing symptoms of psychological disturbance in both clinical (Hofmann et al., 2010; Khoury et al., 2013; Panos et al., 2014) and non-clinical populations (Gotink et al., 2015; Khoury et al., 2015). However, these therapies encounter help-seeking barriers common to face-to-face psychological interventions (Boggs et al., 2014), such as cost, resources, time, stigma, travel and restricted availability (Boetcher et al., 2014).

1.1. mHealth

Cuts to National Health Service funding mean that mental health trusts now have less money to spend on services than they did in 2012 (Royal College of Psychiatrists, 2018). These resource issues have led to a greater focus on proactive, preventative measures aimed at promoting good mental health in non-clinical populations (McDaid et al., 2017), an approach which is thought to be efficacious (Goldie et al., 2016) and cost-effective (Sassi et al., 2015).

Such interventions can be disseminated to multiple people without engulfling the time and resource budgets of professionals (Schueller and Parks, 2014). Hence a focus on positive “self-help” interventions (in which there is little or no therapist-patient contact) with empirically supported effects upon wellbeing (Boiler et al., 2013). People find the idea of “mHealth” (the use of mobile telecommunication technologies for the delivery of health and wellbeing services; Steinhubl et al., 2013; Price et al., 2014) agreeable rather than off-putting (Ryan et al., 2010). Therefore, the increasing ubiquity of smartphone applications (“apps”) presents an opportunity to deliver cost-effective, evidence-based, portable and private self-help services, avoiding many of the aforementioned help-seeking barriers encountered by face-to-face services (Lee and Jung, 2018).

However, despite the large number of self-help apps available, recent reviews found few methodologically robust studies, indicating that the majority of mHealth apps are not supported by evidence-based research (Donker et al., 2013; Payne et al., 2015; Grist et al., 2017).

1.2. Mindfulness apps

Mindfulness apps typically use audio recordings to guide formal meditation practices, in which attention is focused upon the body and the breath. Most apps recommend meditating daily for about 10 min, while sitting still in a quiet place (Roquet and Sas, 2018). Such platforms are widely available (Plaza et al., 2013), and appealing; one survey found adults prefer to access mindfulness interventions online and alone, rather than in a face-to-face group (Wahbeh et al., 2014), and the mindfulness app industry is “booming” (Perez, 2018).

Yet the efficacy of these apps is uncertain. In a review of the quality of 700 iPhone mindfulness apps, Mani et al. (2015) found that just 4% provided mindfulness training and education, and only one was supported by empirical evidence. Furthermore, a meta-analysis of mindfulness-based self-help apps by Plaza et al. (2013) found that only two apps met the inclusion criteria, neither of which was supported by empirical evidence.

However, a small number of individual studies do offer some support for the benefits of mindfulness-based self-help apps. The current literature focuses overwhelmingly on “Headspace”, which was rated as the highest-quality mindfulness app by Mani et al. (2015). The effectiveness of Headspace is supported by multiple RCTs (Bennike et al., 2017; Economides et al., 2018; Howells et al., 2016; Noone and Hogan, 2018; Rosen et al., 2018; Yang et al., 2018), and uncontrolled studies (Laurie and Blandford, 2016; Taylor et al., 2016; Wen et al., 2017; Wylde et al., 2017).

1.3. Attrition and daily use

One commonly cited limitation of these studies is the high number of participants who chose not to complete the assigned mindfulness practice. Recent meta-analyses found the average attrition rate from mindfulness-based self-help interventions to be 37% (Cavanagh et al., 2014), which is higher than that of non-mindfulness self-help mHealth interventions (31%; Melville et al., 2010).

However, this problem of attrition could be overstated; it is grounded in the “universalist” (Farias and Wilholm, 2016, p.329) assumption that daily mindfulness practice is necessary for benefits to be realised (Grossman et al., 2004; Kabat-Zinn, 1990; Segal et al., 2002). A recent meta-analysis of 43 MBCT and MBSR studies (Parsons et al., 2017) found that on average participants completed just 64% of assigned “homework” practice. This figure is likely to overestimate adherence, because it only includes data from course completers. The authors therefore suggested that future studies should also record and report the adherence levels of non-completers.

Some evidence suggests that participants’ level of adherence to assigned mindfulness practice correlates with gains in wellbeing and trait mindfulness (Carmody and Baer, 2008), and reductions in risk of depression (Crane et al., 2014; Hawley et al., 2014). However, these studies used MBCT and MBSR rather than apps, and correlational findings do not tell us about the direction of the association. Rather than daily practice leading to positive outcomes, it could be that positive outcomes motivate daily practice, with those who feel any benefits more reason to continue with the activity than those who do not feel any benefits. Or, since perhaps 70% of people use mindfulness practice intermittently (Brady and Draper, 2019), as a “quick fix” (Wyatt et al., 2014), it is possible that they drop out of research studies early because once they feel better they see no reason in continuing.

Therefore, studies that recommend daily mindfulness practice may lack ecological validity, because they typically require participants to practise mindfulness in an artificial, regimented manner. Furthermore, most studies ignore the possibility that quality of practice may be more important than quantity (Del Re et al., 2013), and rely on subjective self-reporting of adherence, which can lead to overestimations of mindfulness practice time when compared to objective measures (Wahbeh et al., 2011). Indeed, contradictory evidence suggests that daily practice is not necessary for positive outcomes. In 13 of the 24 mindfulness studies reviewed by Vetere et al. (2009), there was no association between practice and outcome, a finding Plaza García et al.
(2017) replicated using a mindfulness app.

1.4. The present study

The present study aims to determine whether the use of a particular mindfulness app is associated with positive outcomes, and whether daily practice is necessary for these outcomes to be realised. Since Headspace is currently overrepresented in the literature, we evaluated an alternative self-help mindfulness-based app: “Calm” (Calm.com, 2018). Calm has recently overtaken Headspace in popularity (Weissman, 2017; Perez, 2018); it has been downloaded 26 million times, with 50,000 new users each day (Sawers, 2018). However, Calm appears in just one non-refereed case study, in which the app was used in conjunction with an exercise program to assist an individual’s recovery from a physical injury (Hagerty, 2017). The study was not focused on mental wellbeing, and the design made it difficult to delineate the app’s contribution to recovery.

Although Calm offers a paid subscription service, we utilised the free, introductory “7 Days of Calm” course, comprised of seven daily 10 minute sessions. Users must complete each session before unlocking the next, and the app encourages users to complete one session each day. The current literature suggests that the use of mindfulness-based self-help apps may lead to positive outcomes. Yet it is unclear whether daily practice is necessary for positive outcomes to be realised. Therefore, we proposed the following hypothesis: Both daily and intermittent mindfulness practice will be associated with significant gains in wellbeing, trait mindfulness and self-efficacy.

A recent review by Tomlinson et al. (2018) noted that the quantitative literature on trait mindfulness relies heavily on self-report measures, which can lead to significant response biases. They argue that the field will benefit from research that also applies qualitative methods. Therefore, we employed a mixed methods approach, aiming to produce both an objective measure of outcomes, and novel insights into participants’ subjective experiences of the app.

Qualitative methods seek to understand phenomena through the eyes of those studied (Bryman, 2015), acknowledging that participants are experts in their own experiences (Kvale, 1996), capable of providing rich, detailed and unexpected data. There is little qualitative research relevant to the present study. A meta-analysis of 15 qualitative studies exploring participants’ experiences of face-to-face MBCT and MBSR courses (Wyatt et al., 2014) found common barriers to practice to be time commitment and an inability to grasp the core concepts of mindfulness. A recent study by Banerjee et al. (2017) identified similar themes. However, these findings may not translate to self-help apps. Laurie and Blandford (2016) is the only known qualitative study exploring experiences of a self-help mindfulness app (Headspace). Time constraints were again cited as the primary barrier, but negative perceptions of mindfulness and low self-efficacy also arose as interesting themes. However, there were only 16 participants, and the necessity of daily practice was not discussed.

The present study is therefore the first mixed methods evaluation of a self-help mindfulness app within a student population, the qualitative aspect of which explores the following research question: What were participants’ subjective experiences of the app?

2. Method

2.1. Participants

Of the 419 individuals who initially expressed an interest, 269 students from a Scottish university (81% female, 19% male) participated in the first stage of the study, with a mean age of 23.89 (SD = 7.23), comprised of 49 first-years, 45 second-years, 49 third-years, 54 fourth-years and 72 postgraduates. The sample was predominantly White Caucasian (84% white, 16% non-white), with 192 native English speakers and 77 non-native speakers. From this sample, 133 individuals participated in the second stage of the study (see Fig. 1). Exclusion criteria constituted those currently diagnosed with a mental health condition, those without a personal smartphone, and those with over 60 h of meditation experience. Both the quantitative and qualitative analyses used the same sample of participants at each stage. Participants were recruited through email and on-campus flyers. The only incentive was the opportunity to partake in a “free online mindfulness course”. Participants were informed of their right to withdraw from the study at any time. This study was conducted without randomisation or a control group.

2.2. Materials

2.2.1. Five Facet Mindfulness Questionnaire - Short Form (FFMQ-SF; Bohlmeijer et al., 2011)

The FFMQ-SF is a 24-item version of the original 39-item FFMQ (Baer et al., 2006), which measures trait mindfulness. Each item is a statement, responded to on a five-point Likert scale. It contains both positively and negatively worded items. The FFMQ-SF has strong psychometric properties which are comparable to those of the original FFMQ (Bohlmeijer et al., 2011). Although the FFMQ-SF consists of five subscales (observing, describing, acting with awareness, non-judging of inner experience, and non-reactivity to inner experience), we followed other researchers (De Bruin et al., 2016; Schellekens et al., 2017) in using the average of all FFMQ-SF item scores, with higher scores indicating greater mindfulness. Cronbach’s α for the present sample was

---

Fig. 1. Flow of participants through each stage of the study, detailing the amount and cause of participant attrition at each stage. This flowchart adheres to CONSORT (Consolidated Standards of Reporting Trials) guidelines (Schulz et al., 2010).
0.83.

2.2.2. Generalized Self-Efficacy scale (GSE; Schwarzer and Jerusalem, 1995)

The GSE is a 10-item measure of perceived self-efficacy. Each item is a statement, responded to on a four-point Likert scale. It contains only positively worded items. Higher average scores indicate a greater level of self-efficacy. The GSE was chosen because of its brevity and strong psychometric properties (Scholz et al., 2002). Cronbach’s α for the present sample was 0.87.

2.2.3. Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS; Stewart-Brown et al., 2009)

The SWEMWBS is a 7-item version of the original 14-item Warwick–Edinburgh Mental Well-being Scale (WEMWS; Tennant et al., 2007) which measures a combination of eudaimonic and hedonic wellbeing. Each item is a statement about an individual’s experiences over the past two weeks, responded to on a five-point Likert scale. It contains only positively worded items, and a greater average score indicates greater mental wellbeing. The SWEMWBS was chosen because it is brief, widely used, and has strong psychometric properties (Stewart-Brown et al., 2011; Stewart-Brown, 2013) which apply cross-culturally (Morse, 1991, p.122). This approach is thought to enhance the validity and credibility of research, and provide a more complex understanding of issues of public health and health promotion (Farmer et al., 2006).

Quantitative analyses employed a between-subjects, repeated measures design. Baseline and follow-up measurements of trait mindfulness (DV), self-efficacy (DV) and wellbeing (DV) were the dependent variables. Multiple paired t-tests were conducted, with participants grouped according to their level of course engagement (IVs; maximum, medium, minimal). An a priori power analysis indicated that to find a medium sized effect (d = 0.50) of change in the outcome variables from baseline to follow-up, with a power of 0.8, the recommended sample size for a two-tailed paired t-test was 33. Holm-Bonferroni corrections were used to counteract the problem of multiple comparisons.

Qualitative data were analysed thematically (Braun and Clarke, 2006). A thematic analysis identifies, analyses and reports patterns of meaning (themes) across the dataset, in order to answer the research question. This approach enabled concise analysis of a large amount of textual data. The scope of the research question was made deliberately broad, so as to enable a flexible, deductive analysis in which unexpected themes could arise “bottom-up” from the data. Relevant parts of the data were systematically identified, before being coded into candidate themes. After reviewing these themes for consistency, they were refined, defined and named. Finally, a narrative analysis of the themes was conducted. Pseudonyms were used to maintain participants’ confidentiality and anonymity.

3. Results

3.1. Quantitative results

Results indicated that engagement with the Calm app was associated with significant gains in all dependent variables. Daily practice was not necessary for these gains to be realised; intermittent practice was sufficient.

3.1.1. Descriptive statistics

Descriptive statistics, norms and internal consistencies for all baseline measurements are included in Table 1. Means (M) and standard deviations (SD) at baseline suggest that participants were representative of the general student population with respect to trait mindfulness and self-efficacy. All Cronbach’s alpha values (α) were in the moderate to high range (Kline, 1999). Participants who chose to complete the study reported significantly higher levels of trait mindfulness, self-efficacy and wellbeing at baseline than those who did not engage in Stage Two of the study (see Table 2). This suggests that participants with more negative perceptions of themselves and their abilities were less likely to be motivated to engage with the mindfulness app.

3.1.2. Inferential statistics

It was hypothesised that both daily and intermittent course engagement would be associated with significant gains in wellbeing, trait

<table>
<thead>
<tr>
<th>Measure scale (number of items)</th>
<th>Score range</th>
<th>Score M (SD)</th>
<th>α Norms M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWEMWBS (7)</td>
<td>1–5</td>
<td>2.96 (0.45)</td>
<td>0.80</td>
</tr>
<tr>
<td>FFMQ-SF (24)</td>
<td>1–5</td>
<td>3.05 (0.44)</td>
<td>0.83</td>
</tr>
<tr>
<td>GSE (10)</td>
<td>1–4</td>
<td>2.95 (0.46)</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Note. SWEMWBS = Short Warwick-Edinburgh Mental Wellbeing Scale; FFMQ-SF = Five Facet Mindfulness Questionnaire-Short Form; GSE = Generalized Self-Efficacy Scale. For all scales, higher scores are indicative of more extreme responding in the direction of the construct assessed.

a Based on a sample of UK citizens aged 16–24 (Ng Fat et al., 2017; N = 2683).
b Based on a sample of UK university students (Ali et al., 2017; N = 210).
c Based on a sample of UK university students (Pulford et al., 2005; N = 64).
mindfulness and self-efficacy. Of the original sample of participants (N = 269), 133 provided follow-up measurements of these variables. Participants were separated into three groups: participants who had completed 0, 1 or 2 days of the course ("minimum engagement", n = 18), participants who had completed 3, 4 or 5 days of the course ("medium engagement"; n = 33), and participants who had completed 6 or 7 days of the course ("maximum engagement"; n = 82).

Paired-sample t-tests (two-tailed) were conducted to test the null hypotheses that baseline and follow-up mean scores were equal. Prior to conducting the analysis, the assumption of normally distributed reference scores was examined. This assumption was considered satisfied after one outlier was removed from the minimum engagement group (with a self-efficacy score of 0, which was assumed to be a mistake in data entry). Shapiro-Wilks’ tests were successful for each sub-group.

Results of the t-tests are presented in Table 3. These results rejected all null hypotheses for both the maximum and medium engagement groups, indicating significant increases from baseline to follow-up mean scores for wellbeing, trait mindfulness and self-efficacy, with Cohen’s d indicating small to medium effect sizes, based on Cohen’s (1992) guidelines. For the minimal engagement group, there was not enough evidence to reject the null hypotheses, suggesting that this level of course engagement did not lead to any significant gains. A graphical representation of the mean difference between baseline and follow-up scores, separated by level of engagement, is displayed in Fig. 2. This demonstrates that daily mindfulness practice was not necessary for gains to be realised.

3.2. Qualitative findings

Qualitative data were thematically analysed in order to explore the research question: What were participants’ subjective experiences of the app? Four broad themes were identified: motivations for participation, perceived outcomes, facilitators and barriers to engagement and attitudes towards future use. There was no missing data, because the online questionnaire required participants to provide an answer for every question before completion.

Table 2

<table>
<thead>
<tr>
<th>Measure scale</th>
<th>Retained (n = 133) M (SD)</th>
<th>Lost (n = 136) M (SD)</th>
<th>t</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trait mindfulness (FFMQ-SF)</td>
<td>3.29 (0.44)</td>
<td>3.00 (0.46)</td>
<td>2.654</td>
<td>&lt; 0.001</td>
<td>0.64</td>
</tr>
<tr>
<td>Self-efficacy (GSE)</td>
<td>3.07 (0.45)</td>
<td>2.92 (0.46)</td>
<td>0.655</td>
<td>0.01</td>
<td>0.33</td>
</tr>
<tr>
<td>Wellbeing (SWEMWBS)</td>
<td>3.16 (0.47)</td>
<td>2.92 (0.44)</td>
<td>1.938</td>
<td>&lt; 0.001</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Note. t = t value. d = Cohen’s d.

Table 3

<table>
<thead>
<tr>
<th>Measure scale</th>
<th>Engagement level (n)</th>
<th>Baseline M (SD)</th>
<th>Follow-up M (SD)</th>
<th>Proportion of individual increases</th>
<th>t</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trait mindfulness (FFMQ-SF)</td>
<td>Minimum (17)</td>
<td>3.17 (0.43)</td>
<td>3.26 (0.60)</td>
<td>61%</td>
<td>0.989</td>
<td>0.338</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium (33)</td>
<td>3.02 (0.38)</td>
<td>3.20 (0.39)</td>
<td>70%</td>
<td>3.275</td>
<td>0.010</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>Maximum (82)</td>
<td>3.13 (0.43)</td>
<td>3.34 (0.42)</td>
<td>74%</td>
<td>6.097</td>
<td>&lt; 0.001</td>
<td>0.50</td>
</tr>
<tr>
<td>Self-efficacy (GSE)</td>
<td>Minimum (17)</td>
<td>2.87 (0.40)</td>
<td>2.84 (0.52)</td>
<td>39%</td>
<td>-0.266</td>
<td>0.794</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium (33)</td>
<td>2.90 (0.42)</td>
<td>3.03 (0.39)</td>
<td>58%</td>
<td>2.303</td>
<td>0.028</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>Maximum (82)</td>
<td>3.03 (0.50)</td>
<td>3.13 (0.45)</td>
<td>60%</td>
<td>2.920</td>
<td>0.014</td>
<td>0.21</td>
</tr>
<tr>
<td>Wellbeing (SWEMWBS)</td>
<td>Minimum (17)</td>
<td>2.97 (0.42)</td>
<td>3.08 (0.56)</td>
<td>57%</td>
<td>1.012</td>
<td>0.327</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium (33)</td>
<td>2.94 (0.46)</td>
<td>3.09 (0.43)</td>
<td>61%</td>
<td>2.603</td>
<td>0.028</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>Maximum (82)</td>
<td>3.06 (0.46)</td>
<td>3.25 (0.44)</td>
<td>71%</td>
<td>4.017</td>
<td>&lt; 0.001</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Note. t = t value. d = Cohen’s d. Statistically significant p-values were adjusted using Holm-Bonferroni correction, to counteract the problem of multiple comparisons. The column showing the proportion of individuals whose scores increased from baseline to follow up is included to illustrate the question of whether or not mindfulness meditation works for everyone.

3.2.1. Theme 1: motivations for participation

This theme demonstrates that many students chose to participate in the study because they were interested in mindfulness practice, they expected it to quickly improve their wellbeing, and they used the study itself as a motivating factor.

3.2.1.1. Curiosity. Some participants had preconceptions of mindfulness as “a bit wishy-washy, hippy and not particularly helpful”. For example, Sarah saw mindfulness as “a bit of a popularfad, which makes me doubt whether it is really worthwhile engaging with." However, she admitted feeling "a bit of longing for the ability to relax and to have more presence of mind", reflecting a curious interest...
in mindfulness that was expressed by many.

3.2.1.2. Quick-fix. Sarah’s desire to “relax” and “de-stress” were the most frequently cited motivating factors, which participants often linked to the upcoming examination period. This is nicely captured by Phillip’s point that “I am heading into a stressful study period right before exams and thought I would benefit from daily meditation to help reduce some of the stress”. This suggests that most participants saw mindfulness practice as a short-term, quick fix solution to, or protective factor against, mental pressures, which is at odds with the Buddhist conception of mindfulness as a way of being that is to be cultivated over a lifetime.

3.2.1.3. The study. Most participants were aware of claims regarding the efficacy of mindfulness, and used the study as an excuse to give it a try. For example, Nitya said

“I rarely take the time out to meditate, despite knowing that its beneficial for me. I thought [the app] would provide a structure for meditation. I’m also more likely to do things when there’s a goal (7 days for a study).”

This dichotomy between acknowledging one might benefit from mindfulness, and yet not practising as often as one feels one ought to, was a common admission, with some participants equating it to their approach to “regular exercise”. This lack of motivation was often influenced by participants’ low self-efficacy with regards to mindfulness; they “feel as though [they’re] no good at it”, often “because [they] don’t know if [they’re] doing it right”. Therefore, these participants used their participation in the study as a motivating factor, because the “feeling of accountability” would make them “more likely to try and meditate every day”. This is demonstrated by Nitya’s prediction that the concrete “structure” and “goal” of the course would enhance her self-efficacy regarding mindfulness practice.

3.2.2. Theme 2: perceived outcomes

This theme captures participants’ subjective evaluations of the effect that the course had on their wellbeing.

3.2.2.1. Positive outcomes. Many participants experienced tangible benefits, to the surprise of Jiang, who said that

“it was actually a lot better than I thought. I immediately noticed being more optimistic and being happier with my current situation… I’ve really de-stressed over the last 2 weeks”.

Jiang’s improvements in life satisfaction and positive affect imply increases in both eudaimonic and hedonic wellbeing respectively. Furthermore, for many participants these outcomes were associated with an increased tendency to be mindful in everyday life. For example, Millie “was pleased to see that [she] could connect mindfulness and meditation to the wider aspects of [her] life”. That participants were noticing these effects outside of their formal meditation practice time conveys the potential of the app to increase levels of trait mindfulness.

3.2.2.2. Marginal outcomes. However, Jiang’s and Millie’s experiences contrast with Marco’s, who

“felt more calm and in the present moment for a short while after the meditation, but soon after [reverted] back to being flooded with pointless thoughts and distractions again”.

This suggests that meditation only succeeded in placing Marco in a temporary mindful state. Others reported feeling that the course had “marginal” or no effect on them, like Tracy who said “I have pretty high life satisfaction at the moment and I didn’t feel like I was gaining anything from meditating”. This suggests that for some people, who are already content with their mental health, daily mindfulness practice may feel like a waste of time.

3.2.2.3. Negative outcomes. Furthermore, others found that the app “made [them] feel worse”. For example, Chrissie revealed that

“I am having a shit couple of months and my anxiety level is about to hit the ceiling. So, I was willing to try anything—even something I tried and failed at before—to calm down…but [the app] sort’a made shit worse. I was getting anxious just having to sit still and listen to someone jab at me.”

It appears that despite Chrissie’s perceived low self-efficacy regarding mindfulness (which she had “failed at before”), she hoped that it could relieve her distress. Yet it seems that Chrissie misconstrued mindfulness as something that is done “at” her, rather than as an active process. Other participants also displayed a fundamental misunderstanding of the practice as a passive exercise; something that “would magically make me much more focused”. Thus, perhaps less people would have had a negative experience if the app had better explained the conceptual basis of mindfulness.

3.2.3. Theme 3: app features

This theme suggests that some people found it easier to engage with the course than others because they reacted differently to features of the app. Fitting mindfulness practice into one’s daily routine was necessary for most course completers. Some said that features of the app facilitated this, for example Gerald found the app’s “daily reminders” helpful, and Maria found it “easier to keep to [the] habit when you see your record of sessions and so on”. Many apps, including Calm, utilise game-like elements (which can increase engagement and enhance learning in students; Pechenkina et al., 2017) such as progression scores, “streak” rewards for consistent use, and home-screen prompts.

However, this gamification seems to have induced feelings of low self-efficacy in other participants; Gemma observed that “having the Calm App sat there waiting for me stressed me out more, because I felt like I should be doing it” and Alan found breaking his streak made it “hard to get back into it as I felt like I ‘failed’ my commitment to do it in 7 consecutive days”. This suggests the gamification of health-promoting apps may lead users to identify as winners or losers, thus influencing perceived self-efficacy.

Other features of the app also acted as barriers to engagement. While some participants found the voice guiding the sessions to be “soothing”, others complained that it was “irritating” and “not calming enough for me to continue”. Thus, engagement may have been improved with an option to choose a different instructor to deliver the content. Furthermore, others were irritated by the frequency and structure of this verbal guidance: “as soon as I relaxed, the voice would return and take me away from my concentration of my breath. I did like hearing the insights, but the timing bothered me on several occasions”.

These points indicate that the Calm app is not suitably adaptable to the characteristics of individual users. While Cassie may have benefited from more guidance (see Theme 2: perceived outcomes), others wanted “more periods of quiet”. Similarly, while some said they “love” the background sounds, others did not; Chris was “annoyed by the background noise…why is that necessary? Why not absolutely silence?”

Chris seems unaware that one of Calm’s few customisable features was the option to choose different background sounds (birds chirping, flowing water) or to turn them off altogether. Therefore, both the non-customisable aspects of the app and its insufficiently clear user interface led some participants to report negative experiences.

3.2.4. Theme 4: attitudes towards future use

This theme conveys the variation in participants’ opinions about the need to practise mindfulness daily. Some practised daily because they quickly noticed a benefit and “wanted to keep feeling that way”. However, others found themselves “dreading taking the time each day to do it”, equating practice with a “chores”; a tedious task that they felt obligated to complete. This suggests that some people continued with the daily practice despite not feeling like it was a worthwhile use of
their time, perhaps because they feel they made a commitment to the research, or to themselves (see Theme 1: motivations for participation).

For Katherine, this “mandatory” daily meditation “made the process feel more like work rather than a fun thing to try. From now on I will meditate on my own time and enjoy it more.” Like Katherine, many plan to use mindfulness practice “only when [they] feel like it” in the future. Harriett explains that this is because “meditating at my own timing/pace is more effective than meditating due to external (perhaps arbitrary) time pressures”. These responses suggest that there may be a different optimal dose of mindfulness practice for different people. Thus, Harriett may be correct to describe the mental “pressure” of daily practice as an “arbitrary” command that prevents some people from benefitting from the app.

4. Discussion

Most mindfulness-based self-help smartphone apps are not supported by empirical evidence (Mani et al., 2015; Plaza et al., 2013). We aimed to determine whether use of the Calm app was associated with positive outcomes, and whether daily practice was necessary for these outcomes to be realised. To do this, we tested the following hypothesis: Both daily and intermittent mindfulness practice will be associated with significant gains in wellbeing, trait mindfulness and self-efficacy. Results supported the hypothesis, indicating that those who used the Calm app both daily and intermittently reported significant gains, while those who chose not to engage with Calm or only did so briefly did not report significant gains.

We also employed thematic analysis to investigate the research question: What were participants’ subjective experiences of the app? Four themes were identified: motivations for participation, perceived outcomes, app features, and attitudes towards future use. These themes indicated that participants’ experience of the course varied widely, with some experiencing great benefits, and others finding it made them feel worse. Participants suggested that their preferences for either daily or intermittent mindfulness practice, and for various features of the app, may have influenced the extent to which they engaged with and benefitted from the course.

Two main conclusions can be drawn from these results: i) there is now empirical evidence suggesting that the Calm app works; ii) there are important differences in the way that people experience and engage with mindfulness practice.

4.1. Calm is supported by empirical evidence

This study is the first to demonstrate that use of the Calm app is associated with positive mental health outcomes. Since mindfulness practice increases trait mindfulness (Quaglia et al., 2016), and trait mindfulness is positively correlated with wellbeing (Brown and Ryan, 2003), we can speculate that the mindfulness practice within the “7 Days of Calm” course enhanced wellbeing by facilitating increases in trait mindfulness. Comparing this to similar research with student non-student samples, our study contradicts the findings of a small study by Lomas et al. (2015) and Laurie and Blandford (2016), with some participants reporting being unable to manage negative thoughts that arose during meditation, to the point that it worsened their mental health. Likewise, in the present study Chrissie and others demonstrated that spending time alone with one’s thoughts can be an unpleasant experience.

This phenomenon may be underreported, but it is not new. Shapiro (1992) found that 63% of long-term meditators reported “adverse events”. A recent review found that it was “not uncommon” for mindfulness practice to cause agitation, anxiety, discomfort, and confusion, as well as “potentially” psychosis and the resurfacing of trauma memories (Creswell, 2017). Why do such negative outcomes occur? Lomas et al. (2015) speculate that some people find it difficult to “decentre” or “disidentify” from their thoughts; to view them as transient mental events rather than aspects of their being. Since even MBSR instructors may not be adequately trained to manage individuals’ negative experiences in a face-to-face environment (Dobkin et al., 2012), the inability of self-help mindfulness apps to identify and support such individuals is cause for concern. As Farias and Wikholm (2016, p.330) argue, we should refrain from treating “the potential for adverse effects...as the elephant in the room.”

Going forwards, the potential for mindfulness practice to lead to various outcomes, including negative experiences, must be emphasised by proponents. This would help ensure that users begin mindfulness practice with reasonable expectations, and that “unsuccessful” practice does not cause feelings of stigma and guilt. Furthermore, although it is positive that mindfulness practice and other wellbeing activities are becoming more socially accepted, we must ensure that this positivity is framed in light of the current fragmentary evidence base, rather than acquiescing to hyperbolic media coverage which casts mindfulness practice as a “universal panacea for various types of human deficiencies and ailments” (Van Dam et al., 2018).

4.2. Daily practice should not be recommended

Daily mindfulness practice is recommended by Kabat-Zinn (1990), Segal et al. (2002), Grossman et al. (2004), Parsons et al. (2017), and the developers of apps such as Calm and Headspace. However, since
there is little empirical evidence demonstrating that daily practice is always more beneficial than intermittent use, this recommendation appears to be merely an assumption predicated on traditional Buddhist conventions.

In the present study, we obtained quantitative and qualitative follow-up data from participants who did not complete the course. This was an improvement on previous studies which focused only on course completers (Banerjee et al., 2017; Boggs et al., 2014; Chittaro and Vianello, 2016; Forbes et al., 2018; Van Emmerik et al., 2018). Furthermore, while Lee and Jung (2018) and Van Emmerik et al. (2018) relied on participants to self-report their app use, risking over-estimation of adherence (Wahbeh et al., 2011), we collected objective data from the app, providing a more accurate picture of whether or not participants chose to use the course daily. This enabled us to determine that intermittent mindfulness practice was associated with significant positive gains, supporting findings by Vettese et al. (2009) and Plaza García et al. (2017).

In addition, the attitudes towards future use theme demonstrates that while daily practice was optimal for some participants, many found this “mandatory” and seemingly “arbitrary” requirement reduced their motivation to use the app. Some participants self-perceived “failure” to engage with the app on a daily basis had a demotivating effect on future use. Many other users rejected the advice of mindfulness proponents like Kabat-Zinn (1990), indicating that in the future, they plan to use the app only when they feel like it. This reflects the results of research by Brady and Draper (2019) who found that 70% of mindfulness meditation users in his study chose to practise mindfulness intermittently. Furthermore, some participants going into our study with initially high wellbeing reported little to no benefits from daily practice.

These results suggest that the ideal dosage of mindfulness practice varies by person, and that we must challenge the widely-held assumption that daily practice is desirable. Mindfulness proponents should, as a precaution, not recommend daily practice until its benefits are clearly supported by empirical evidence.

4.2.3. Apps should have more customisable features

The qualitative data contains valuable user feedback on particular details of the Calm app, which strengthen the argument that future mindfulness apps ought to provide adaptable and personalised features.

The sub-theme app features reveal that the game-like aspects of Calm lowered the self-efficacy of those who “failed” to comply with daily practice. Furthermore, mean self-efficacy scores were lowest for those who used the app the least, and low self-efficacy was also a notable post-course complaint of participants in qualitative studies by Boggs et al. (2014) and Laurie and Blandford (2016). Since self-efficacy predicts health related outcomes (Duff, 2010; Holden, 1992), and those with low self-efficacy are likely to be less motivated to act (Ajzen, 1991; Bandura, 1997), positive psychological interventions which aim to enhance wellbeing must ensure that they do not set unrealistic targets which may inadvertently lower users' self-efficacy. Therefore, app developers should rethink the gamification of their platforms. While such gamification may encourage subscriptions and increase revenue, the wellbeing of users of a wellbeing app should take priority over developer's financial profit. Since daily practice works for some but not others, in-app targets and congratulation messages should be customisable on the basis of user preference.

Another controversial aspect of the app was the voice, which some found “irritating” and others “soothing”. This echoes research suggesting that voice preference varies (Fraccurro et al., 2010; Kreiman et al., 1992). For example, different people prefer different accents (Souza et al., 2013), and one’s voice preferences can vary depending on one's physiological state (Feinberg et al., 2008). Therefore, to maximise engagement, app developers should enable users to choose from a variety of voices to guide their meditation practice.

Opinion also varied over the right balance of spoken guidance to silence. Guidance helped keep some from being distracted, but others found these interruptions inhibiting and yearned for “more periods of quiet” so that they could fully focus on their own body and mind. These differences may be because participants received exactly the same level of guidance, despite variance in their ability (trait mindfulness). In educational psychology, the “scaffolding” theory holds that we learn most effectively when the level of guidance fits with one's current level of competency (Wood et al., 1976). This theory is supported by a recent study in which participants initially rated both guided and unguided meditation practices favourably, but preferred the unguided practices after six weeks of use (Waelde et al., 2017). Therefore, future apps might give users the option of manipulating the frequency of vocal prompts, with an algorithm removing aspects of the full audio recording as required.

As demonstrated in the negative outcomes sub-theme, some participants failed to adequately grasp the conceptual basis of mindfulness practice, in particular that it is an active rather than a passive exercise. That the efficacious MBSR and MBCT courses devote considerable time to teaching the theoretical underpinnings of mindfulness suggests that it would be good practice for mindfulness apps to do so too. The “7 Days of Calm” course does provide a brief conceptual explanation at the beginning of sessions, but outcomes may have improved if this was reinforced, perhaps at the end of each session. A textual explanation of the concepts behind mindfulness could also be made available within future apps.

Finally, while Calm did enable customisation (by allowing users to turn off or change the background noise) this was not readily apparent to some participants (such as Chris), suggesting that an insufficiently clear user interface was also a barrier to engagement.

4.2.4. We must re-evaluate why people practise mindfulness

The results of the present study suggest that the optimum dose of mindfulness practice varies. This may indicate that different people practise mindfulness for different reasons. Many mindfulness proponents warn that secular Western approaches to mindfulness practice are “diluted...watered down” versions (Dimidjian and Linehan, 2003) that foster a “significant misunderstanding” which “overlooks much of the richness and uniqueness” of Buddhist practice (Walsh and Shapiro, 2006). They argue that the goals of mindfulness are fundamentally spiritual (Dimidjian and Linehan, 2003), and that it should be “a regular lifelong practice in which [trait] mindfulness is cultivated over a period of many years” (Lynam and Baer, 2009).

However, rather than being motivated by these long-term spiritual goals, the motivations for participation theme suggests that almost all participants in the present study (and in Laurie and Blandford, 2016) had different priorities. They saw mindfulness practice as a “quick fix” – a temporary tool for avoiding stress and achieving a state of relaxation in the short-term. In other words, they were using mindfulness in order to enjoy temporary pleasant experiences, a goal which appears anti-pathetic to “dukkha” (Teasdale and Chaskalson, 2011) and the Buddhist reasons for mindfulness practice.

Does this mean that participants practised mindfulness incorrectly? Or should we simply recognise that the Western approach to mindfulness is different, underlined by a short-term pragmatic concern with “what works” (Goldstein, 2002), rather than long-term spiritual goals? Since Westerners and Buddhists have different motivations for practising mindfulness, perhaps the current discourse conflates two fundamentally different forms of mindfulness, one old and one new (Schmidt, 2011). The current literature does not adequately reflect these different motivations for mindfulness practice, and when it does a value judgment is often made in favour of those who pursue longer-term goals. If public health bodies are to continue to promote mindfulness as a short-term self-help tool for managing one’s mental health, this attitude should change. It should both be recognised that mindfulness practice can serve different purposes for different people, and emphasised that it does not work for everyone.
4.3. Limitations

We aimed to determine whether the Calm app works. However, the observational nature of the design of the study precludes a definitive answer to this question. Resource limitations meant that the use of a control group and randomisation was unfeasible. It is therefore possible that the reported positive outcomes were caused by a placebo effect, rather than the Calm app. Indeed, the theme motivations for participation indicated that many participants expected to benefit from the course, especially in the short-term. This reflects previous qualitative research, which found that participants expected mindfulness to be a “quick fix” for their mental wellbeing (Wyllatt et al., 2014), and were “determined that this was going to be very beneficial” (Williams et al., 2011, p.387). Such high outcome expectancy increases the motivation to perform an activity (Ajzen, 1991), which increases the chances of a placebo effect (Geers et al., 2005). To control for these possibilities, future studies should adopt a randomly sampled, active-control design as in Noone and Hogan (2018).

However, even if these results were obtained using an RCT, it would still be possible that participants’ reported positive outcomes did not result from their use of Calm. They could instead be due to “regression to the mean”; a statistical phenomenon in which more extreme baseline scores tend to be followed by more moderate follow-up scores. Since this occurs even in the absence of an intervention (Linden, 2013), we must be wary of concluding that the “7 Days of Calm” course caused the observed outcomes; they merely indicate that this may be the case.

Besides the placebo effect, other variables may have confounded the results. Firstly, although people with over 60 h of meditation experience were excluded from the study, there potentially remained considerable variation in baseline trait mindfulness among participants. This may have influenced wellbeing outcomes; Lilja et al. (2013) found that more experienced meditators scored higher on the “observing” facet of the FFMQ than less experienced meditators, and Bravo et al. (2016) theorise that those who score highly on all mindfulness facets are most likely to report positive wellbeing. Secondly, participants may have varied in their ability to operate a smartphone, making app use easier for some than others. Thirdly, the “7 Days of Calm” course was a brief intervention, and only one follow-up measurement was taken, during what was for many participants an emotionally challenging academic period. Therefore, reported changes in outcomes from baseline to follow-up may have been influenced by factors other than the app, such as participants’ meditation experience, technological abilities, and exam performance.

Another limitation was the reliance on subjective self-report measures, which may be inaccurate (Tomlinson et al., 2018). For example, those high in trait mindfulness might have a greater capacity to notice their mind wandering, thereby rating themselves more conservatively on the FFMQ-SF than those who are oblivious to their mind wandering, despite it occurring more frequently. Also, participants who grasped the conceptual basis of mindfulness practice may have been able to determine the “correct” responses to the follow-up FFMQ-SF (Grossman, 2011), and this could have facilitated a demand effect (in which participants report a post-intervention gain because they believe this is what the researchers want them to do). These biases could be mitigated by assessing outcome variables using ecological momentary assessment (EMA), which involves repeated sampling of participants’ experiences, at random times, in their natural environments (Shiffman et al., 2008). EMA has outperformed traditional questionnaire measures of trait mindfulness (Moore et al., 2016). Thus, future studies could collect more ecologically valid quantitative and qualitative data by implementing EMA; it seems ideally suited to smartphone technology, and it could be administered within the same app that is used to deliver mindfulness content.

Finally, it is important to be reflective when undertaking qualitative research (Elliott et al., 1999). The lead author approached this topic with five years of experience of self-help mindfulness practice, using both the Headspace and Calm apps. Furthermore, the question of whether daily practice is important resonates particularly strongly with him. Indeed, whilst writing this paper he went through phases of both daily and intermittent practice, and remains ambivalent about what works best for him.

5. Conclusion

This study is the first to evaluate the influence of the Calm app upon the mental health of its millions of users. The results indicate that Calm works, because many participants who used it reported significant and tangible gains in wellbeing, trait mindfulness and self-efficacy. Furthermore, the results demonstrate that intermittent practice works for, and appeals to, many people (which challenges the ecological validity of much previous research). Moreover, this study challenges the conventional claim that mindfulness must be practised daily in order for benefits to be realised.

Since the findings show that daily practice can have a negative effect upon mental wellbeing, and that daily practice may not benefit those who are already high in wellbeing, mindfulness proponents and app developers should not claim that daily practice is optimal. They should also increase awareness of the potential for mindfulness practice to result in negative outcomes, resist the idea that mindfulness practice works for everyone, and re-evaluate why Westerners choose to practise mindfulness. In light of this, future mindfulness apps should provide more opportunities for personalisation and customisation. In particular, they should explain the conceptual basis of mindfulness practice to beginners, who may otherwise feel like failures for being unable to achieve a mindful state after an initial ten-minute session.

This study has contributed to a growing body of literature which suggests that personalised, evidence-based self-help mindfulness apps have the potential to both enhance and diminish users’ wellbeing. It indicates that:

- The Calm app works.
- Intermittent mindfulness practice works, and is widely used.
- Mindfulness practice can have negative consequences.
- There may be no benefit to mindfulness practice when one’s well-being is already very high.
- Beginners need to be educated about the conceptual basis of mindfulness.
- Users of mindfulness-based apps would benefit from more customisable features.

Declaration of competing interest

None.

References


Shapiro, S.L., 1992. A preliminary study of long term meditators: goals, efficacy a...