Meta-Analytic Evidence for Effects of Mindfulness Training on Dimensions of Self-Reported Dispositional Mindfulness

Jordan T. Quaglia, Sarah E. Braun, Sara P. Freeman, Michael A. McDaniel, and Kirk Warren Brown
Virginia Commonwealth University

Improvements in stable, or dispositional, mindfulness are often assumed to accrue from mindfulness training and to account for many of its beneficial effects. However, research examining these assumptions has produced mixed findings, and the relation between dispositional mindfulness and mindfulness training is actively debated. A comprehensive meta-analysis was conducted on randomized controlled trials (RCTs) of mindfulness training published from 2003–2014 to investigate whether (a) different self-reported mindfulness scale dimensions change as a result of mindfulness training, (b) key aspects of study design (e.g., control condition type, population type, and intervention type) moderate training-related changes in dispositional mindfulness scale dimensions, and (c) changes in mindfulness scale dimensions are associated with beneficial changes in mental health outcomes. Scales from widely used dispositional mindfulness measures were combined into 5 categories for analysis: Attention, Description, Nonjudgment, Nonreactivity, and Observation. A total of 88 studies (n = 5,787) were included. Changes in scale dimensions of mindfulness from pre to post mindfulness training produced mean difference effect sizes ranging from small to moderate (g = 0.28–0.49). Consistent with the theorized role of improvements in mindfulness in training outcomes, changes in dispositional mindfulness scale dimensions were moderately correlated with beneficial intervention outcomes (r = 0.27–0.30), except for the Observation dimension (r = 0.16). Overall, moderation analyses revealed inconsistent results, and limitations of moderator analyses suggest important directions for future research. We discuss how the findings can inform the next generation of mindfulness assessment.

Keywords: meditation, meta-analysis, mindfulness training, mindfulness-based stress reduction (MBSR), trait mindfulness

Supplemental materials: http://dx.doi.org/10.1037/pas0000268.supp

Scientific interest in mindfulness has never been greater. Since the early 2000s there has been an exponential increase in research publications on the topic (K. W. Brown, Creswell, & Ryan, 2015), extending into mindfulness theory and conceptualization, basic science, and applied science. A considerable body of research has examined the effectiveness of mindfulness-based interventions for a variety of mental health, physical health, and associated neurobiological outcomes. Recent meta-analyses (e.g., Hofmann, Sawyer, Witt, & Oh, 2010; Khoury et al., 2013; Sedlmeier et al., 2012) have shown that improvements in a number of mental health symptoms, in particular, are associated with participation in such mindfulness-based interventions as mindfulness-based stress reduction (MBSR; Kabat-Zinn, 1990) and mindfulness-based cognitive therapy (Segal, Williams, & Teasdale, 2002). Mindfulness training has likewise been incorporated into a wide range of additional interventions, supplementing existing intervention content in an effort to maximize benefits. Yet it is currently unknown what specific psychological processes are responsible for any training effects.

Enhancements in mindfulness, or mindfulness training skills, are thought to be the central active ingredients that produce intervention benefits, and over the past decade numerous efforts have been made to measure mindfulness and skills theorized to be associated with it. To date these efforts have almost exclusively focused on the development and use of self-report instruments, and they have been widely used in mindfulness intervention research. Thus, studies using these measures can provide empirical information to indicate (a) which psychological processes change over the course of mindfulness training, (b) whether there are reliable moderators of such changes, and (c) whether such changes predict outcomes targeted by mindfulness interventions. The intent of the present meta-analytic review was to examine these three central questions, focusing on measures of trait or dispositional mindfulness for adults, which have seen the most widespread application in research.

Theoretical and Practical Value of Review

Classical Buddhist scholarly accounts of mindfulness highlight a close, clear-minded attention to, or awareness of, what is per-
received in the present (e.g., Análayo, 2003; Bodhi, 2011), and present-oriented attention or awareness are features highlighted in many accounts and measures of mindfulness (see review by Quaglia, Brown, Lindsay, Creswell, & Goodman, 2015). However, there is no one meaning of mindfulness, and no single authoritative account that trumps all others (Análayo, 2013; Dreyfus, 2011); all definitions of mindfulness are rooted in particular scholastic and practice traditions and must be understood from within those contexts. Thus, we can speak of classical mindfulness, nondual mindfulness, clinical mindfulness, and so on (Análayo, 2013; Bodhi, 2011; Dunne, 2011).

Dispositional, or trait, measures of mindfulness are similarly diverse in their theoretical origin. Many of the scales were derived from clinical conceptions of mindfulness. For example, the Philadelphia Mindfulness Scale (PHLMS; Cardaciotto, Herbert, Forman, Moitra, & Farrow, 2008) takes as its conceptual guide Kabat-Zinn’s (e.g., 1994) description of mindfulness. Other scales, such as the Freiburg Mindfulness Inventory (Walach, Buchheld, Buttenmüller, Kleinknecht, & Schmidt, 2006) and the Mindful Attention Awareness Scale (MAAS; K. W. Brown & Ryan, 2003) are considerably influenced by classical scholarly conceptions of mindfulness. The Kentucky Inventory of Mindfulness Skills (KIMS; Baer, Smith, & Allen, 2004) and the Five Facet Mindfulness Questionnaire (FFMQ; Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006) largely developed their operationalizations from dialectical behavior therapy (DBT; Linehan, 1993) but have been widely used in mindfulness intervention research. Importantly, the FFMQ was derived from a factor analysis including five published measures of dispositional mindfulness, so there is considerable overlap between subscales of the FFMQ and other (sub)scales, especially the MAAS and KIMS.

In part because of this theoretical diversity, the scales also vary in what phenomena are assessed. Table 1 presents the basic psychometric features of each of the eight widely known dispositional scales developed for adult respondents (Quaglia et al., 2015), and the number of randomized controlled studies in which each was used in the 12-year period of this review (2003–2014) to assess training-related changes in dispositional mindfulness scale scores. All the scales aim to measure quality of attention, and as already noted, this is considered central to mindfulness in both scholarly and most clinical conceptions of mindfulness. For example, mindfulness has been described as: “An alert but receptive equanimous watching” (Análayo, 2003, p. 60); “Watchfulness, the lucid awareness of each event that presents itself on the successive occasions of experience” (Bodhi, 2011, p. 21); “Paying attention in a particular way: on purpose, in the present moment, and nonjudgmentally” (Kabat-Zinn, 1994, p. 4).

The MAAS focuses exclusively on this quality of attentiveness. Other scales include factors derived from particular theoretical traditions or clinical models. For example, the PHLMS includes both an awareness factor and an acceptance factor, consistent with Kabat-Zinn’s (1994) description of mindfulness. Consistent with the DBT model of therapeutic change, the KIMS and FFMQ include factors assessing acting with awareness, nonjudgment of experience, observation of experience, description or labeling of internal experience, and (on the FFMQ only) nonreactivity. Thus, the scales differ in the constructs assessed. As a result, this review had both theoretical and practical aims through an assessment of whether the various subscales or dimensions tapped by different measures are sensitive to mindfulness training-related change, as their origin conceptualizations of mindfulness claim they should be. Due to the variety of dimensions evident across scales, we

Table 1

<table>
<thead>
<tr>
<th>Scale name</th>
<th>No. of items</th>
<th>Factor(s)</th>
<th>Reliability</th>
<th>Validity</th>
<th>No. of RCTs using scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive and Affective Mindfulness Scale--Revised (CAMS-R)</td>
<td>12</td>
<td>Attention, Present-Focus, Awareness, Acceptance</td>
<td>Internal consistency</td>
<td>Convergent, discriminant, concurrent</td>
<td>4</td>
</tr>
<tr>
<td>Five Facet Mindfulness Questionnaire (FFMQ)</td>
<td>39</td>
<td>Observe, Describe, Act With Awareness, Nonjudge, Nonreact Mindfulness</td>
<td>Internal consistency</td>
<td>Convergent, discriminant, concurrent</td>
<td>33</td>
</tr>
<tr>
<td>Freiburg Mindfulness Inventory (FMI)</td>
<td>14</td>
<td>Observe, Describe, Act With Awareness, Accept Without Judgment</td>
<td>Internal consistency, test–retest</td>
<td>Convergent, discriminant, concurrent, predictive</td>
<td>12</td>
</tr>
<tr>
<td>Kentucky Inventory of Mindfulness Skills (KIMS)</td>
<td>39</td>
<td>Observe, Describe, Act With Awareness, Accept Without Judgment</td>
<td>Internal consistency, test–retest</td>
<td>Convergent, discriminant, concurrent, predictive</td>
<td>7</td>
</tr>
<tr>
<td>Mindful Attention Awareness Scale (MAAS)</td>
<td>15</td>
<td>Attention/Awareness</td>
<td>Internal consistency, test–retest, parallel forms</td>
<td>Convergent, discriminant, concurrent, predictive, incremental</td>
<td>44</td>
</tr>
<tr>
<td>Philadelphia Mindfulness Scale (PHLMS)</td>
<td>20</td>
<td>Awareness, Acceptance</td>
<td>Internal consistency</td>
<td>Convergent, discriminant, concurrent, predictable, incremental</td>
<td>4</td>
</tr>
<tr>
<td>Southampton Mindfulness Questionnaire (SMQ)</td>
<td>16</td>
<td>Decentered Awareness, Letting Go of Reacting, Accepting, Opening Awareness to Difficult Experience Curiosity, Decentering</td>
<td>Internal consistency</td>
<td>Convergent, discriminant, incremental</td>
<td>2</td>
</tr>
<tr>
<td>Trait Toronto Mindfulness Scale (TMS)</td>
<td>13</td>
<td></td>
<td>Internal consistency</td>
<td>Convergent, discriminant, incremental</td>
<td>4</td>
</tr>
</tbody>
</table>

Note. Scales were published as follows: CAMS-R (Feldman, Hayes, Kumar, Greeson, & Laurenceau, 2007); FFMQ (Baer et al., 2006); FMI (Walach et al., 2006); KIMS (Baer et al., 2004); MAAS (Brown & Ryan, 2003); PHLMS (Cardaciotto et al., 2008); SMQ (Chadwick et al., 2008); and TMS (Davis, Lau, & Cairns, 2009). The number of RCTs using each scale refers to located studies that examined training-related changes in mindfulness scale scores.
chose to focus specifically on dimensions most commonly and consistently assessed.

As the seminal meta-analysis addressing how mindfulness scale dimensions may be differentially affected by interventions, we first aimed to address what purported mindfulness scale dimensions show significant improvements over the course of mindfulness training and intervention. Sedlmeier et al.’s (2012) meta-analysis of meditation-based training outcomes found that across studies, mindfulness training produced a moderately sized increase in “mindfulness” (mean $r = .34$), but this outcome category did not differentiate between mindfulness measures, subscales, or dimensions, and included such psychological constructs as dissociation, concentration, and curiosity, along with dimensions more commonly associated with descriptions of mindfulness (e.g., present-centered attention). Similarly, a meta-analysis by Visted, Vollstad, Nielsen, and Nielsen (2015) found a moderate increase in mean dispositional mindfulness over the course of training ($g = 0.53$), without differentiating dimensions of the construct. Further, Visted et al.’s review included studies published up to March, 2011 only, and a considerable amount of relevant research has been published since that time. The present comprehensive analysis of studies available through 2014 aimed to examine training-related changes in the specific dimensions that have been associated with various operationalizations of mindfulness.

The second aim of this review was to ask whether there are reliable moderators of intervention-based changes in mindfulness. This question has been little explored in individual studies, even though it is apparent that studies of mindfulness interventions vary in important ways that could have measurable consequences (Khoury et al., 2013; Sedlmeier et al., 2012; Visted et al., 2015). We therefore identify and address study design characteristics (e.g., wait-list vs. active controls), research participant populations (e.g., clinical vs. nonclinical), and features of mindfulness intervention delivery (e.g., intervention length) to evaluate via meta-analysis. This investigation informs whether intervention-based changes in mindfulness are conditional upon such key factors, which is relevant for both mindfulness theory and application.

If meaningfully significant intervention-related changes in dispositional mindfulness scale dimensions are found, such changes do not in themselves support the claim that mindfulness scale dimensions constitute these interventions’ active ingredients. Therefore, the third aim of this review was to examine whether changes in dispositional mindfulness scale scores were correlated with treatment outcomes. Mindfulness has been theorized to promote numerous beneficial outcomes (K. W. Brown, Creswell, & Ryan, 2007; Vago & Silbersweig, 2012), including mental health and well-being, emotion regulation, behavior regulation, physical health, and positive interpersonal outcomes. The strongest test of whether mindfulness dimensions are active ingredients of mindfulness training requires tests of mediation. Yet to date, very few articles assess whether mindfulness predicts or is related to beneficial changes across mental health and well-being outcomes, and even fewer include statistical tests of mediation. In this meta-analytic review, we therefore leverage the available studies that report associations between changes in mindfulness scale scores and changes in mental health outcome variables to provide initial evidence addressing whether mindfulness can help to explain the beneficial effects of mindfulness interventions.

This analysis provides an objective evaluation of scale performance, and thereby can serve both to inform theory on the active psychological processes in mindfulness training and facilitate the selection of mindfulness measures in future research. Importantly, this analysis is not intended to examine the construct validity of the scales; whether or how well each scale actually measures mindfulness is a theoretical and empirical matter beyond the scope of this review. Yet knowing which scale dimensions and, by implication, which purported mindfulness phenomena are most affected by mindfulness interventions may be valuable for the further development of mindfulness assessments. The current scales represent first-generation efforts to operationalize mindfulness and there is some controversy regarding their fidelity to theoretical understandings of mindfulness (e.g., Grossman, 2011). Identification of well-performing scales or subscales, in combination with well-specified theory, can focus future psychometric development efforts through the selection of items comprising those scales or subscales. The development of objective, behavioral assessments of mindfulness can also be aided through this analysis, by targeting behaviors for assessment that are tapped by the most effective self-report measures.

Method

Search Procedure and Study Selection

Prior research has demonstrated the potential for nonrandomized studies to artificially inflate effect size estimates (Higgins & Green, 2008). Therefore, only randomized controlled trials (RCTs) were included in the present review. We searched ERIC, PsycINFO, ProQuest, PubMed, Scopus, and Web of Science to locate all relevant articles published between 2003 and 2014 inclusive. Mindfulness and specific scale names (including their abbreviations) were used as search keywords. When options were available in the database, we used more specific strategies: in PsycINFO, we filtered search results to include only RCTs, and then selected publications according to their use of one or more mindfulness scales; in PubMed, we selected articles that cited the initial scale validation studies. For articles published since 2010, we cross-checked our publications list with that of the Mindfulness Monthly Newsletter (see Black, 2010), a research newsletter that tracks and organizes citations for newly published mindfulness science literature. We supplemented our search by checking the references of retrieved articles. The title and abstract of each identified article was initially screened for the use of a mindfulness scale in the context of a mindfulness intervention. If the study clearly met these criteria or information was insufficient to make a determination, we retrieved the published report. Additional inclusion and exclusion criteria (see above) were applied to all retrieved publications. Every effort was made to obtain information needed to compute effect sizes, including requests from corresponding authors when necessary information was lacking in the published report. In addition, a request for unpublished studies on the topic was posted in relevant electronic discussion forums (“Mindfulness” and “MBSR” listservs). As such, the review includes both published and thesis/dissertation studies, as well as studies obtained through electronic requests for unpublished studies.
Inclusion and Exclusion Criteria

In the initial study selection process, we considered studies including any of the eight trait mindfulness scales designed for adult respondents that are shown in Table 1. To maximize the statistical power of our analyses, and to best assess specific dimensions associated with trait mindfulness, we grouped the mindfulness (sub)scales into the following categories, based on a cutoff of $r > .70$ (see correlations in the online supplemental materials Table S1): Attention (MAAS and FFMQ Act With Awareness), Description (FFMQ and KIMS Describe), Nonjudgment (FFMQ Nonjudgment, KIMS Accept, and PHLMS Accept), Nonreactivity (FFMQ Nonreactivity), and Observation (FFMQ Observe, KIMS Observe, and PHLMS Awareness). We treated (sub)scales as distinct dimensions when their correlations were lower than $r = .70$ (see Table S1). Table 2 provides sample items from each of these empirically distinct dimensions, the (sub)scales that were grouped into that dimension, and available correlations between (sub)scales. Insufficient correlational evidence supported inclusion of the CAMS-R, FMI, SMQ, and TMS dispositional measures in analyses of these dimensions; in addition, a small number of intervention studies used these scales. Thus we omitted them from analyses.

Studies were included in the analysis if they (a) reported pre-intervention and post-intervention scores on the scale(s) used; (b) focused on adult populations that have received the most study to date—namely college students and healthy community members (both considered normative populations), and adults with physical or mental health conditions (clinical populations); (c) were published between April 2003, when the first well-validated mindfulness scale was published, and December 2014; and (d) were published in the English language. For analysis of relations between changes in mindfulness scale scores and intervention outcomes, an additional criterion was applied: (e) one or more correlations between the scale(s) and pre–postintervention mental health outcome(s) of a subjective (e.g., self-reported) or objective (e.g., germane task performance, neural activation) nature.

To advance understanding by comparing common means through which mindfulness training is delivered, we incorporated a range of interventions involving mindfulness, from interventions focused solely on increasing mindfulness to those that integrate mindfulness training alongside nonmindfulness components. Studies examining both clinical and normative populations were included, as mindfulness is widely believed to enhance well-being among healthy individuals and to reduce symptoms among those suffering from mental or physical health conditions (Khoury et al., 2013). We included randomized design studies with either active or inactive control groups in the present review. Although active control groups typically offer the most rigorous tests of psychological interventions, the majority of RCTs on mindfulness training to date have examined its effects relative to wait-list controls. We therefore also included wait-list controlled studies, which substantially increased the number of studies to be assessed. Although we excluded case studies, no additional restrictions were placed on minimal sample size.

Potential Moderators

The mindfulness intervention literature is diverse in the populations studied, as well as the variety of study designs and types of mindfulness interventions used. Thus we also tested a number of moderators of the effects of interest. Specifically, we examined whether pre–post changes in self-reported mindfulness differed according to a number of categorical moderators. To account for potential differences in value of mindfulness interventions for those seeking treatment for specific conditions from those engaging in mindfulness interventions for other reasons, we considered Population (normative or clinical) as a moderator. Control Condition Type (active or inactive) was also considered, because prior research indicates that effect sizes for mindfulness or other meditation-based interventions differ between active and inactive controls (Khoury et al., 2013; Sedlmeier et al., 2012). Within the category of mindfulness interventions one also finds considerable differences in exactly how mindfulness instruction is delivered. Therefore, we examined two moderators pertaining to these differences at the broadest levels of overall approach to providing mindfulness instruction and the intervention duration. In detail, Intervention Type included three categories: mindfulness-integrated, which include a mindfulness component alongside other, nonmindfulness components (e.g., cognitive therapy); mindfulness-based, which incorporate heterogeneous mindfulness practices; and specific mindfulness meditation, which focus primarily on the training of one form of mindfulness (e.g., focused attention).

Table 2

<table>
<thead>
<tr>
<th>Mindfulness scale dimension</th>
<th>Sample item</th>
<th>Distinct (sub)scales and reported correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attention</td>
<td>It seems I am “running on automatic” without much awareness of what I’m doing.</td>
<td>FFMQ Act Aware, MAAS ($r = .744$)</td>
</tr>
<tr>
<td>Description</td>
<td>I’m good at finding words to describe my feelings.</td>
<td>KIMS Describe, FFMQ Describe ($r = .988$)</td>
</tr>
<tr>
<td>Observation</td>
<td>When I’m walking, I deliberately notice the sensations of my body moving.</td>
<td>KIMS Observe, FFMQ Observe ($r = .967$)</td>
</tr>
<tr>
<td>Nonjudgment</td>
<td>I criticize myself for having irrational or inappropriate emotions.</td>
<td>KIMS Accept, FFMQ Nonjudge ($r = .955$)</td>
</tr>
<tr>
<td>Nonreactivity</td>
<td>In difficult situations, I can pause without immediately reacting.</td>
<td>FFMQ Nonreact</td>
</tr>
</tbody>
</table>

Note. Correlations from Baer (2015), Brown (2015), and Cardaciotto et al. (2008). FFMQ = Five Facet Mindfulness Questionnaire (Baer et al., 2006); KIMS = Kentucky Inventory of Mindfulness Skills (Baer et al., 2004); MAAS = Mindful Attention Awareness Scale (Brown & Ryan, 2003); PHLMS = Philadelphia Mindfulness Scale (Cardaciotto et al., 2008).
Intervention length was also tested as a moderator. Due to widely varying reporting methods and ways of quantifying intervention length, we delineated length of intervention into three broad categories; as multiple studies reported seven sessions for mindfulness interventions (e.g., Aikens et al., 2014; J. M. Davis, Manley, Goldberg, Smith, & Jorenby, 2014), seven served as the best cutoff for distinguishing longer term interventions: <7 training sessions; 7+ sessions without retreat day or other supplementary component; 7+ sessions with retreat or other supplementary component. A recent meta-analysis (Visted et al., 2015) identified the presence or absence of a retreat component (e.g., half-day or full-day practice in mindfulness) as a significant moderator of effect size. We broadened this category to include all forms of supplementary training (e.g., one-on-one phone coaching). In preliminary metaregression analyses, we assessed two additional moderators specifically pertinent to publication bias, namely sample size and year of publication. A significant correlation between effect size and sample size can indicate publication bias (Egger, Davey Smith, Schneider, & Minder, 1997), and year of publication affords a test of time-lag publication bias (Ioannidis & Trikalinos, 2005). Online supplemental material Table S2 specifies the moderators examined here.

Coding Procedures

Two project supervisors (doctoral students trained in meta-analytic techniques) and one research assistant confirmed study relevance and coded the study details, including type of mindfulness intervention, study design, mindfulness scale(s) used, outcome measures, participant descriptions, and the statistical results for the mindfulness scales used in each study. The researchers/assistant coded a set of 10 preliminary articles to ensure high intercoder reliability before completing the coding used in analyses. Once coding was complete, the two project supervisors jointly reviewed and corrected any typographic or substantive errors in the coding. Remaining disagreements were resolved via email consultation with the corresponding author of the study in question.

Statistical Methods

Effect size considerations. For articles reporting more than one effect size (e.g., Ainsworth, Eddershaw, Meron, Baldwin, & Garner, 2013) the coder first determined if the two effect sizes stemmed from independent samples (i.e., separate studies). If this was the case, then both effect sizes were included in the analysis. However, if the multiple effect sizes stemmed from the same sample (e.g., Oken et al., 2010) additional steps were required to determine which effect size would be used.

Specifically, for studies that included a comparison between a single treatment condition and multiple control conditions, we chose the comparison between treatment and the more rigorous, active control condition. Due to the small proportion of included studies that reported statistical results from one or more follow-up assessments, we used the effect size that compared the baseline and the first postintervention assessment. Some studies used two or more mindfulness scales to assess the same construct (e.g., MAAS and FFMQ Act with Awareness; Morone, Rollman, Moore, Li, & Weiner, 2009), and due to potential problems resulting from multiple administrations of the same items on these scales, we chose to prioritize the scale most consistent with the overall measurement approach (e.g., FFMQ Act Aware over MAAS when additional FFMQ subscales were used).

Data considerations and model choice. We used Comprehensive Meta-Analysis (CMA; version 3.3.07) software and the Hedges and Olkin (1985) method to analyze the data. All analyses used a random effects model and weighted studies by the inverse of the sampling error variance. Because some samples had relatively small sample sizes, we used Hedges’s g (Hedges, 1981) to estimate standardized mean differences. We first meta-analyzed the effect sizes pertaining to effects of mindfulness training on changes in dispositional mindfulness scale scores. In separate meta-analyses, we tested whether these changes correlated with changes in one or more outcome variables. Analyses also tested for potential publication/availability bias among the sample, and moderation effects using standard procedures available in the CMA software. Regarding publication bias analytic methods, we used multiple advanced methods for assessing publication bias, wherein the available literature fails to represent the total population of studies in a systematic way (Rothstein, Sutton, & Borenstein, 2005). These methods were metagression of germane moderators (Stanley & Jarrell, 1989), funnel plots (Sterne & Egger, 2005), and Duval and Tweedie’s (2000) trim and fill analysis.

Results

From Search Results to Included Studies

The included studies represented a heterogeneous set of mindfulness interventions, populations, and control condition types, consistent with the wide application of mindfulness training. Results from the literature search, including the number of studies meeting inclusion criteria, are presented in Figure 1. Of the 1,015 germane publications and unpublished studies for which the full text was retrieved, 88 met all remaining inclusion and exclusion criteria. In the final analyses, these 88 studies were included to assess the effect of mindfulness training on self-reported mindfulness, and 14 of these studies contained relevant results that associated mindfulness scale changes with changes in other germane dependent variables.

Publication Bias

As an initial test of publication bias, we examined the potential moderating roles of both sample size and year of publication using metagression (Stanley & Jarrell, 1989). Metagression results did not reveal systematic bias, as neither variable predicted effect sizes concerning intervention-related changes in mindfulness scale dimensions. Further sensitivity analyses were also performed, in which publication bias would be suggested by an asymmetrical distribution in the funnel plot of effect sizes of one or more mindfulness scale dimensions. The funnel plot for the Attention dimension is presented in the online supplemental material Figure S1, wherein six studies were imputed to correct for asymmetry. This indicates that, for Attention, stud-
ies with small sample sizes and high effect sizes were disproportionately represented over studies with small samples and low effect sizes. Results from Duval and Tweedie’s (2000) trim and fill analysis are reported in each table. When present, asymmetry is consistent with the inference that small effect sizes are suppressed in the reviewed literature. Although Attention was the only dimension for which funnel plots and trim and fill analyses revealed publication bias when all studies were included, publication bias was also evident in either the funnel plot or trim and fill analyses for specific moderation analyses to be reported below.

Figure 1. Flow diagram for study selection in meta-analyses. RCT = randomized controlled trial; CAMS-R = Cognitive and Affective Mindfulness Scale-Revised; FMI = Freiburg Mindfulness Inventory; SMQ = Southampton Mindfulness Questionnaire; TMS = Trait Toronto Mindfulness Scale.

Mindfulness Intervention-Related Changes in Mindfulness Scale Dimensions

Tables 3–7 present detailed results for each of the five dispositional mindfulness scale dimensions examined. Each weighted mean effect size is accompanied by its 95% confidence interval (CI), an $I^2$ statistic indicating degree of heterogeneity (i.e.,

1 Consistent with the publication bias literature, the word suppression refers to studies that are not readily available (e.g., unpublished studies). Here the word does not imply deceit.
### Table 3
Meta-Analysis and Sensitivity Analyses for Attention Dimension and Modifiers

<table>
<thead>
<tr>
<th>Subdistribution</th>
<th>N</th>
<th>k</th>
<th>Mean ES</th>
<th>95% CI</th>
<th>( I^2 )</th>
<th>Leave one out</th>
<th>Trim &amp; fill</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ESmin ESmax</td>
<td>k_imputed ESAdj ΔES Q ( \rho )</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active</td>
<td>1,546</td>
<td>29</td>
<td>.207</td>
<td>[.108, .306]</td>
<td>.00</td>
<td>.196 .219</td>
<td>0 0</td>
</tr>
<tr>
<td>Inactive</td>
<td>3,625</td>
<td>50</td>
<td>.536</td>
<td>[.440, .672]</td>
<td>65.33</td>
<td>.522 .571</td>
<td>13 .378 .178</td>
</tr>
<tr>
<td>Participant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical</td>
<td>2,212</td>
<td>34</td>
<td>.486</td>
<td>[.360, .613]</td>
<td>51.78</td>
<td>.457 .505</td>
<td>0 0</td>
</tr>
<tr>
<td>Not clinical</td>
<td>2,959</td>
<td>45</td>
<td>.397</td>
<td>[.283, .511]</td>
<td>56.31</td>
<td>.353 .412</td>
<td>9 .264 .133</td>
</tr>
<tr>
<td>Intervention type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mindfulness-integrated</td>
<td>237</td>
<td>5</td>
<td>.448</td>
<td>[.193, .704]</td>
<td>.00</td>
<td>.366 .588</td>
<td>1 .427 .021</td>
</tr>
<tr>
<td>Mindfulness-based</td>
<td>4,172</td>
<td>60</td>
<td>.447</td>
<td>[.341, .553]</td>
<td>63.49</td>
<td>.420 .458</td>
<td>5 .372 .075</td>
</tr>
<tr>
<td>Single mindfulness meditation</td>
<td>762</td>
<td>14</td>
<td>.404</td>
<td>[.269, .540]</td>
<td>6.86</td>
<td>.326 .428</td>
<td>0 0</td>
</tr>
<tr>
<td>&lt;7 training sessions</td>
<td>1,039</td>
<td>21</td>
<td>.341</td>
<td>[.188, .494]</td>
<td>41.35</td>
<td>.304 .359</td>
<td>0 0</td>
</tr>
<tr>
<td>7+ training sessions without retreat</td>
<td>2,776</td>
<td>38</td>
<td>.492</td>
<td>[.381, .604]</td>
<td>49.14</td>
<td>.465 .517</td>
<td>2 .458 .034</td>
</tr>
<tr>
<td>7+ training sessions + retreat/other</td>
<td>1,356</td>
<td>20</td>
<td>.461</td>
<td>[.249, .674]</td>
<td>70.31</td>
<td>.320 .496</td>
<td>7 .193 .268</td>
</tr>
</tbody>
</table>

Note. \( N \) = total sample size from all included samples; \( k \) = number of samples (i.e., number of effect sizes); Mean ES = weighted mean observed effect size; 95% CI = 95% confidence interval; \( I^2 \) = percent of variance not attributed to random sampling error; \( ES_{\text{min}} \) = lowest effect size after removing one study at a time; \( ES_{\text{max}} \) = highest effect size after removing one study at a time; \( k_{\text{imputed}} \) = number of trim and fill imputed effect sizes; \( ES_{\text{Adj}} \) = trim and fill adjusted observed mean; \( \Delta ES \) = change in effect size after trim and fill adjustment; \( Q (\rho) \) = statistical value and significance for mean difference in effect size between moderator subgroups.

Variance not attributed to random sampling error, and results from two different types of sensitivity analyses: leave one out, which assesses how much the overall effect changes due to removing a single study (Patsopoulos, Evangelou, & Ioannidis, 2008), and trim and fill, which estimates the influence of any potential missing studies on effect size estimates (Duval & Tweedie, 2000). Finally, these tables include between-groups \( Q \) tests for significant differences in effect size between moderator

### Table 4
Meta-Analysis and Sensitivity Analyses for Description Dimension and Modifiers

<table>
<thead>
<tr>
<th>Subdistribution</th>
<th>N</th>
<th>k</th>
<th>Mean ES</th>
<th>95% CI</th>
<th>( I^2 )</th>
<th>Leave one out</th>
<th>Trim &amp; fill</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ESmin ESmax</td>
<td>k_imputed ESAdj ΔES Q ( \rho )</td>
</tr>
<tr>
<td>Total</td>
<td>2,727</td>
<td>34</td>
<td>.275</td>
<td>[.231, .320]</td>
<td>.00</td>
<td>.263 .283</td>
<td>0 0</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active</td>
<td>807</td>
<td>10</td>
<td>.269</td>
<td>[.218, .320]</td>
<td>.00</td>
<td>.175 .273</td>
<td>0 0</td>
</tr>
<tr>
<td>Inactive</td>
<td>1,920</td>
<td>24</td>
<td>.294</td>
<td>[.204, .384]</td>
<td>.00</td>
<td>.267 .332</td>
<td>0 0</td>
</tr>
<tr>
<td>Participant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical</td>
<td>1,543</td>
<td>17</td>
<td>.286</td>
<td>[.237, .334]</td>
<td>.00</td>
<td>.278 .302</td>
<td>0 0</td>
</tr>
<tr>
<td>Not clinical</td>
<td>1,184</td>
<td>17</td>
<td>.216</td>
<td>[.102, .332]</td>
<td>.00</td>
<td>.188 .260</td>
<td>4 .155 .061</td>
</tr>
<tr>
<td>Intervention type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mindfulness-integrated</td>
<td>625</td>
<td>4</td>
<td>.408</td>
<td>[.150, .667]</td>
<td>53.41</td>
<td>.360 .492</td>
<td>1 .392 .016</td>
</tr>
<tr>
<td>Mindfulness-based</td>
<td>1,757</td>
<td>26</td>
<td>.266</td>
<td>[.219, .314]</td>
<td>.00</td>
<td>.218 .275</td>
<td>0 0</td>
</tr>
<tr>
<td>Single mindfulness meditation</td>
<td>345</td>
<td>4</td>
<td>.202</td>
<td>[.009, .412]</td>
<td>.00</td>
<td>.176 .241</td>
<td>1 .175 .027</td>
</tr>
<tr>
<td>&lt;7 training sessions</td>
<td>303</td>
<td>5</td>
<td>.132</td>
<td>[−.091, .355]</td>
<td>.00</td>
<td>.115 .153</td>
<td>0 0</td>
</tr>
<tr>
<td>7+ training sessions without retreat</td>
<td>1,171</td>
<td>15</td>
<td>.301</td>
<td>[.251, .350]</td>
<td>.00</td>
<td>.293 .397</td>
<td>2 .295 .006</td>
</tr>
<tr>
<td>7+ training sessions + retreat/other</td>
<td>1,253</td>
<td>14</td>
<td>.185</td>
<td>[.074, .295]</td>
<td>.00</td>
<td>.169 .217</td>
<td>3 .155 .030</td>
</tr>
</tbody>
</table>

Note. \( N \) = total sample size from all included samples; \( k \) = number of samples (i.e., number of effect sizes); Mean ES = weighted mean observed effect size; 95% CI = 95% confidence interval; \( I^2 \) = percent of variance not attributed to random sampling error; \( ES_{\text{min}} \) = lowest effect size after removing one study at a time; \( ES_{\text{max}} \) = highest effect size after removing one study at a time; \( k_{\text{imputed}} \) = number of trim and fill imputed effect sizes; \( ES_{\text{Adj}} \) = trim and fill adjusted observed mean; \( \Delta ES \) = change in effect size after trim and fill adjustment; \( Q (\rho) \) = statistical value and significance for mean difference in effect size between moderator subgroups.
subgroups. These dimensions appear to be differentially affected by mindfulness training. The effect size for mindfulness training-related change on Attention, the scale category with the largest number of studies, was moderate in size ($g = 0.44, CI [0.35, 0.52]$). Effect sizes for three out of four other mindfulness scale dimensions were also moderate (cf. Cohen, 1992): 0.44, CI [0.33, 0.54] (Nonjudgment), 0.49, CI [0.36, 0.64] (Nonreactivity), and 0.47, CI [0.37, 0.58] (Observation). The effect size for Description was comparatively small ($g = 0.28, CI [0.23, 0.32]$). Mean effect size estimates appear reasonably
constant when considering their 95% CIs and results from two distinct sensitivity analyses (leave one out; trim and fill).

**Moderators of Change in Mindfulness Scale Dimensions**

A metaregression (Stanley & Jarrell, 1989) was conducted to identify potential moderators of the effect sizes pertaining to intervention-related changes in mindfulness scale dimensions. For each of the five mindfulness scale dimension categories (Attention, Description, Observation, Nonjudgment, and Nonreactivity), metaregressions performed using CMA assessed the following potential moderator variables: intervention type, intervention length, participant type, and control condition type. The online supplemental material Table S3 summarizes the metaregression results. As a conservative cutoff that would retain potentially important moderators, we decided to select moderators if they had $\beta$ weights greater than ±.20 for two or more mindfulness scale dimensions. All four moderators met this criterion, and were thus retained for further moderation analyses.

**Moderation by control condition type.** We next compared the changes in mindfulness scale dimensions according to the two broad types of control conditions used in the studies examined (see second section of Tables 3–7). Overall, there were substantially more studies comparing mindfulness interventions with inactive (i.e., wait-list; $k = 50$) than active ($k = 29$) control conditions. Only Attention had more than 15 ($k = 29$) studies with active control conditions, limiting the interpretability of effect size estimates for active control condition designs for the other 4 dimensions. A between-groups $Q$ test revealed that there was a significant difference in effect size estimates for changes in Attention, such that effect sizes were lower for active controls ($g = 0.21$, CI [0.11, 0.31]) than for wait-list control conditions ($g = 0.56$, CI [0.44, 0.67]).

2 Non-significant effect size differences in the same direction between Control Type subgroups were apparent for the other 4 mindfulness scale dimensions (effect sizes ranged from 0.27 to 0.42 for active controls and from 0.29 to 0.55 for waitlist controls). However, there was not a relative decrease in $I^2$ evident from all studies to $I^2$ based on moderator subgroups; this suggests that additional heterogeneity within some moderator subgroups has yet to be explained.

3 **Moderation by participant type.** We also compared the change in mindfulness scale dimensions obtained using normative and clinical research participant populations (see third section of Tables 3–7). There were more observations for normative ($k = 45$) than for clinical ($k = 34$) populations. Between-groups $Q$ tests revealed that there were no significant differences in participant type for any of the scale dimensions. However, the relative size of the effect of mindfulness training on mindfulness scale scores varied; the effect sizes for clinical populations exceeded the estimates of normative populations for the Attention, Description, and Nonjudgment dimensions (effect size estimates ranged from 0.29 to 0.38) for normative populations (effect size estimates ranged from 0.21 to 0.27).
to 0.50 for clinical samples, and from 0.22 to 0.39 for normative samples), whereas normative estimates were higher for Nonreactivity ($g = 0.49, CI [0.34, 0.64]$) and Observation ($g = 0.52, CI [0.39, 0.65]$) than clinical sample estimates ($g = 0.45, CI [0.22, 0.69]$ and $g = 0.40, CI [0.22, 0.59]$), respectively. In addition to nonsignificant differences between participant type subgroups, there were also no decreases in $I^2$ from all studies to $I^2$ for Population Type subgroups, suggesting participant type may not account for heterogeneity in effect of mindfulness training.

**Moderation by intervention type.** We similarly examined the magnitude of changes in mindfulness scale dimensions resulting from the three types of interventions. The largest number of studies used mindfulness-based interventions ($k = 60$), with a considerably smaller number of both mindfulness-integrated interventions ($k = 5$), and specific mindfulness meditation trainings ($k = 14$). The relative magnitude of effect size estimates for each type of intervention differed across scale categories, though between-groups $Q$ tests did not reveal significant differences between them. Mindfulness-based interventions had the largest mean effect size for Observation (Table 7; $g = 0.52, CI [0.39, 0.65]$), whereas effect sizes for Attention, Description, Nonjudgment, and Nonreactivity (Tables 3–6) were largest for mindfulness-integrated interventions ($g = 0.45, CI [0.19, 0.70]; g = 0.41, CI [0.15, 0.67]; g = 0.71, CI [0.36, 1.06]; and $g = 0.85, CI [0.48, 1.23]$, respectively).

**Moderation by intervention length.** The largest number of studies used interventions of 7 + training sessions without an additional component, such as a half- or full-day retreat ($k = 38$), with a smaller number of studies using <7 training sessions ($k = 21$), and 7 + sessions with an additional component (e.g., retreat day; $k = 20$). The relative magnitude of effect size estimates for each intervention length differed across scale categories, but for all mindfulness scale dimensions, the effect size estimates for interventions without an additional component exceeded estimates for interventions with such a component or for briefer interventions. For Nonjudgment and Nonreactivity (Tables 5–6), a between-groups $Q$ test revealed that there was a significantly larger effect size between interventions of 7 + sessions without an additional component ($g = 0.57, CI [0.38, 0.75]; g = 0.73, CI [0.54, 0.92]$, respectively) and the other intervention lengths, and this same comparison was marginally significant for Description (Table 4; $g = 0.30, CI [0.25, 0.35]$ for 7 + sessions without an additional component, and $g = 0.13, CI [-0.09, 0.36]$ and $g = 0.19, CI [0.07, 0.29]$ for <7 training sessions and 7 + training sessions with an additional component, respectively). Between group $Q$ tests indicated no significant differences in effect sizes according to intervention length for Attention nor Observation.

### Assessing Relations Between Mindfulness Scale Dimension Changes and Intervention Outcome Changes

To test the claim that mindfulness is a process variable that supports mindfulness intervention outcomes, the relations between changes in mindfulness scale dimensions and changes in outcome variables pertaining to mental health and well-being (e.g., depression, anxiety, quality of life) was examined. As indicated earlier, only a subset of studies ($k = 14$) examined these relations. To maximize the statistical power of these analyses, we assessed the overall relation between changes in a given mindfulness scale dimension and beneficial changes in mental health measures. For each study, we averaged all reported effect sizes relating pre–post change in a mindfulness scale with change scores in mental health outcome variables. We coded all effect sizes such that positive correlations indicated beneficial change. Results from this meta-analysis are presented in Table 8, with average correlations across seven or more studies for each mindfulness scale, along with their 95% CI, an $I^2$ statistic indicating degree of heterogeneity (i.e., variation attributed to the presence of moderators), and results from sensitivity analyses (leave one out; trim and fill). Average correlations for Attention ($r = 0.29, CI [0.21, 0.38]$), Description ($r = 0.27, CI [0.16, 0.38]$), Nonjudgment ($r = 0.17, CI [0.07, 0.26]$), and Nonreactivity ($r = 0.30, CI [0.15, 0.43]$) mindfulness scale dimensions with beneficial changes in outcome variables were moderate in magnitude, whereas the average correlation for Observation was small ($r = 0.16, CI [0.09, 0.24]$). The lowest point of the 95% CI for Attention ($r = 0.21$) remained moderate in size, providing the strongest evidence of dispositional mindfulness correlating with beneficial changes in mental health, whereas the lowest values for CIs across all other scale dimensions were small.

### Table 8

**Meta-Analysis and Sensitivity Analyses for Correlations of Change in Mindfulness Scale Dimensions and Other Outcomes**

<table>
<thead>
<tr>
<th>Subdistribution</th>
<th>$N$</th>
<th>$k$</th>
<th>Mean $r$</th>
<th>95% CI</th>
<th>$I^2$</th>
<th>Leave one out</th>
<th>Sensitivity analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$ES_{min}$</td>
<td>$ES_{max}$</td>
</tr>
<tr>
<td>Attention</td>
<td>1,005</td>
<td>14</td>
<td>.297</td>
<td>[.214, .376]</td>
<td>44.81</td>
<td>.567</td>
<td>.313</td>
</tr>
<tr>
<td>Description</td>
<td>475</td>
<td>7</td>
<td>.272</td>
<td>[.159, .379]</td>
<td>34.83</td>
<td>.223</td>
<td>.303</td>
</tr>
<tr>
<td>Nonjudgment</td>
<td>678</td>
<td>9</td>
<td>.266</td>
<td>[.172, .354]</td>
<td>33.55</td>
<td>.232</td>
<td>.293</td>
</tr>
<tr>
<td>Nonreactivity</td>
<td>475</td>
<td>7</td>
<td>.300</td>
<td>[.162, .426]</td>
<td>55.26</td>
<td>.241</td>
<td>.334</td>
</tr>
<tr>
<td>Observation</td>
<td>678</td>
<td>9</td>
<td>.162</td>
<td>[.086, .236]</td>
<td>0.0</td>
<td>.152</td>
<td>.176</td>
</tr>
</tbody>
</table>

**Note.** All correlations were coded such that positive values indicate beneficial change. $N$ = total sample size from all included samples; $k =$ number of samples (i.e., number of effect sizes); Mean $ES =$ weighted mean observed effect size; 95% CI = 95% confidence interval; $I^2 =$ percent of variance not attributed to random sampling error; $ES_{min} =$ lowest effect size after removing one study at a time; $ES_{max} =$ highest effect size after removing one study at a time; $k_{trimmed} =$ number of trim and fill imputed effect sizes; $ES_{adj} =$ trim and fill adjusted observed mean; $ME_{ES} =$ change in effect size after trim and fill adjustment; $Q (p) =$ statistical value and significance for mean difference in effect size between moderator subgroups.
in size. Sensitivity analyses revealed evidence for some potential publication bias, especially as indicated by trim and fill nonzero values for number of imputed studies and corresponding changes in effect size. The small number of studies in each meta-analysis limits interpretability of effect size estimates and sensitivity analyses.

Discussion

The availability of new tools can drive theoretical and empirical developments in psychological science. Measures of mindfulness have been in research use for over a decade. Yet to date, little empirical evidence has been available to show what psychological processes tapped by the measures are responsive to mindfulness training, and whether other intervention outcomes are associated with those training-related changes in mindfulness scale dimensions. The present meta-analytic review sought to address these unknowns via three questions: whether there are intervention-related changes in five purported dimensions of mindfulness assessed by widely used self-report scales; whether there are important moderators of these changes; and finally, whether such changes are related to pre- to postintervention changes in mental health outcomes. The findings of this review can inform about the measurement of mindfulness in psychological science to date, and help to spur theoretical and empirical developments in mindfulness assessment.

In response to our first question, meta-analyses found that mindfulness training affected most of the purported dimensions of dispositional mindfulness as measured by the KIMS, FFMQ, MAAS, and PHILMS. Specifically, four of the five dimensions assessed by these measures - Attention, Nonjudgment, Nonreactivity, and Observation demonstrated sensitivity to mindfulness training, as indicated by moderate mean effect sizes. The effect size estimate for changes in Description was considerably lower than the other dimensions. We found some evidence for publication bias in primary meta-analyses of this review of 88 RCTs. Specifically, the funnel plot and trim and fill analysis for the Attention dimension imputed six studies consistent with an inference that small effect sizes are suppressed in this literature. However, the change in magnitude of effect size after accounting for these imputed studies was minimal, suggesting little publication bias. More generally, examination of 95% CIs revealed they were narrow in range and did not cross zero, suggesting that total mean effect size estimates accurately represented the mean effect of mindfulness training on dispositional mindfulness dimensions.

Regarding our second question about moderators of training-related changes in trait mindfulness, analyses revealed that the effect size estimate of change in Attention was significantly larger for studies with inactive control conditions than for those using active control conditions. Though the effect size differences between inactive and active controls were not statistically significant for the other mindfulness scale dimensions (Q tests), estimates for inactive control conditions were consistently higher as well. This suggests an overestimation of the true effects of mindfulness training on dimensions of dispositional mindfulness scales in wait-list controlled studies. An important question is whether nonspecific factors (e.g., demand characteristics) account for larger score increases in wait-list controlled studies, or if factors other than mindfulness training per se promote these scale score changes in active control interventions. In addition, effect size estimates were highest for interventions with greater than seven training sessions that did not have an additional component (e.g., retreat day), in contrast to similar analyses conducted previously on total mindfulness scale scores (Visted et al., 2015). These results suggest that the inclusion of a retreat day or other additional training component may not be critical to the effects of interventions on the dispositional mindfulness scale dimensions assessed by the included scales. Yet this conclusion was based on a rather crude categorization of training length, so must be qualified. Further research is required to determine what training length and composition of training components offers maximal training impact, and for which normative and clinical populations.

That many commonly studied moderators (e.g., clinical vs. nonclinical participants, intervention length) did not reliably explain variation in effects of training on mindfulness scale dimensions suggests there may be additional moderating variables relevant to mindfulness training that have yet to be widely assessed, or which were not considered in the present review. In other words, lack of significant moderation may indicate the presence of other key factors driving differences in mindfulness scale changes over the course of intervention, such as training facilitator education level, previous experience with the training model, extent of personal experience with mindfulness, participant-level “dosage” of mindfulness, and study quality indicators (e.g., blinding outcome experimenters to conditions). It is also possible that some of these differences could be explained by interactions between moderators or by the ways in which such moderators as intervention type or length were categorized here. There is not yet a standard approach to classifying distinct types or durations of mindfulness training; attention to such potentially important variables would help to guide both research and application. Finally, lack of support for moderation may also be attributed to the small k in some moderator subgroups. Analyses should be repeated as additional data become available. To support these efforts, the data included in the present meta-analyses are available in the Supplemental Materials.

An alternative explanation for the significant effects of mindfulness training on changes in the scale dimensions examined here is that demand characteristics are responsible for change in scale scores (Grossman, 2011). This is an important consideration for any future mindfulness assessment development effort. Although the demand explanation is plausible, particularly when examining pre- to postintervention changes, there are reasons to question it. First, a number of self-report mindfulness measures do not assess the construct in face valid ways. In other words, respondents do not necessarily know that mindfulness is being assessed by the scales, and those examined here have been found relatively impervious to social desirability (K. W. Brown & Ryan, 2003; Baer et al., 2006; Baer et al., 2004). Thus, response bias is not a robust explanation for the scale score changes observed here.

Further, and pertinent to addressing the third question posed in this review, meta-analysis found sizable relations between changes in four of the five mindfulness scale dimensions and changes in generally well-validated, and in many cases, commonly used outcome measures. Unless response bias can be assumed on these measures as well, then the observed relations likely reflect valid responses on the mindfulness scales. Yet the fact that only a minority of the RCT studies examined here reported associations between changes in mindfulness scale responses and outcome
measures supports a provisional conclusion regarding the role of the purported mindfulness scale dimensions in affecting intervention outcomes. Clearly more research should focus on examining the relations between changes in self-reported mindfulness scale dimensions and other intervention outcomes. Further, more studies involving mediation analyses are needed to conclusively determine whether improvements in mindfulness dimensions are indeed the active ingredients of mindfulness training; assessing change in mindfulness prior to outcome assessment would be more conclusive. Despite this limitation, these meta-analytic results provide the strongest evidence to date that dimensions assessed by commonly used mindfulness scales may represent active ingredients of change in mental health outcomes examined in research on several widely studied mindfulness interventions.

Toward Next-Generation Mindfulness Assessment

Self-report measures of mindfulness have been subject to question, particularly as to whether they tap the dimension(s) of mindfulness understood in the canonical and other expositions of mindfulness (e.g., Grossman, 2011). The intent of the present review was not to establish the construct validity of the various mindfulness measures—that is, to determine whether the measures examined here are reliable and valid measures of mindfulness. Rather we aimed to point future research toward, first, the examination of psychological processes that may be active ingredients of mindfulness trainings and interventions; and second, to help guide development of “second generation” assessments of mindfulness scale dimensions and training skills, including measures of a non-self-report nature (e.g., Frewen et al., 2014; Levinson, Stoll, Kindy, Merry, & Davidson, 2014). The present review suggests several dimensions and skills that are sensitive to change over relatively brief periods of time and that may be important drivers of training success. However, the extent self-report measures likely assess mindfulness and mindfulness-related skills in a very basic sense only (K. W. Brown, Ryan, Loverich, Biegel, & West, 2011) and the potential for psychometric improvement seems high. Meta-analytic evidence provided herein, namely training-related changes in dispositional mindfulness scale scores and their association with mental health outcomes, suggests that efforts to improve measurement of mindfulness may facilitate the understanding of mindfulness training processes.

Limitations and Future Directions

This meta-analytic review was limited in several ways that can inform about important directions for future research. A particularly important limitation was that only a subset of dispositional mindfulness scales met inclusion criteria to warrant a meta-analytic investigation. Even for the scales included, the comparatively small number of studies in some categories for moderator analyses demands that the moderation results in general, though especially for intervention type and length, be interpreted with caution. Especially for those analyses with a relatively small number of studies, differences between average effect sizes may therefore provide better estimates of actual population differences in effect size magnitude than statistical tests of significance (Hedges & Olkin, 1985). That stated, the Attention dimension had the largest number of studies (more than twice the number of any other dimension), so effect size estimates of change from the primary and moderation analyses of this dimension may be more reliable than those from the other scale dimensions. Certainly more research is needed to provide better estimates of these dimensions and their relations to intervention outcomes, as well as effect size estimates of other dimensions tapped by mindfulness scales not included in this review. Moreover, results revealed a discrepancy in the type of studies examining pre–post changes in dispositional mindfulness, with substantially fewer germane mindfulness-integrated and single mindfulness training RCTs.

Finally, this study does not address the stability of the observed changes and relations over time. The posttest measures were generally completed soon after the end of training. The effects of mindfulness training on the outcomes examined here may dissipate with time or instead increase over time, at least for trainees who continue to practice mindfulness. Longitudinal studies would help to address the stability of the effects across time.

Conclusion

To date, the effect of mindfulness interventions on the enhancement of mindfulness has been largely unknown, and the utility of self-report-based mindfulness scales to assess the phenomenon has been a source of theoretical debate. Here we provided meta-analytic evidence from RCT studies that mindfulness training does appear to produce increases in dimensions purported to be either key elements of mindfulness or important mindfulness-related skills developed through training. This review suggests that targets for future research initiatives on the assessment of mindfulness should consider the dimension(s) expected to be trained. Mindfulness training had moderate effects on four of five dimensions assessed by dispositional mindfulness scales. Yet the dimensions that showed the strongest improvements over time varied according to population and intervention moderators, and this variation suggests that different scale dimensions may tap or represent distinct intervention targets. Having presented the strongest support to date that dimensions of self-reported mindfulness are affected by mindfulness training, and that these changes are associated with other beneficial changes, we close with a call for next generation mindfulness measures to advance the investigation of how, for what, and for whom various types of mindfulness training are beneficial.

References

References marked with an asterisk indicate studies included in the meta-analysis.


"Roeser, R. W., Schonert-Reichl, K., Jha, A., Cullen, M., Wallace, L., Wilensky, R., ... Harrison, J. (2013). Mindfulness training and reductions in teacher stress and burnout: Results from two randomized,


Accepted November 10, 2015

Revision received November 9, 2015

Received June 22, 2015

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.