Integrating mindfulness and connection practices into preservice teacher education improves classroom practices

Matthew J. Hirshberg*, Lisa Flook, Robert D. Enright, Richard J. Davidson

University of Wisconsin Madison, USA

ARTICLE INFO

ABSTRACT

Teachers vary in their ability to enact effective teaching practices. We randomly assigned 88 early education preservice teachers to standard teacher education or teacher education plus a 9-week mindfulness-based intervention. Using the Classroom Assessment Scoring System (CLASS) as our primary outcome, we assessed effective teaching practices at baseline and at a 6-month follow-up that occurred during full-time student teaching. Mindfulness, negative affect, and well-being were assessed at baseline, post-test, and follow-up. At follow-up, we observed significant GROUP × time interactions on all major CLASS domains: Instructional supports, Emotional supports, and Classroom organization favoring the intervention group (Cohen's d's 0.53–0.65). Daily mindfulness practice was significantly associated with intervention group improvements on Instructional supports ($r = .39$) and Classroom organization ($r = .38$). No group differences were observed on negative affect or well-being. Implications for teacher education are discussed.

1. Introduction

Effective teachers are essential to successful educational systems, accounting for more variance in student outcomes than any other school-based factor (Rivkin, Hanushek, & Kain, 2005). Highly effective teachers mitigate challenges associated with poverty, improving student graduation rates, college matriculation rates, and labor market outcomes (Chetty, Friedman, & Rockoff, 2014). Educating effective teachers is therefore paramount. But teaching is a complex and emotionally demanding practice (Hargreaves, 1998; Shulman, 1986). Attrition, burnout, and heterogeneity in performance suggest that teachers enter the profession with significant variability in their capacity to enact effective practices (Cheng & Zamarro, 2018).

The problem of adequately preparing new teachers for the complex demands of the profession has long been appreciated by teacher education scholars. Kennedy (1999; 2016) described the issue as the "the problem of enactment", or the gap between what teachers know and are able to do. Graduating new teachers with knowledge of topics such as content, curricula, and pedagogy is undoubtedly important. However, according to the problem of enactment, knowledge alone is not sufficient. In addition to teaching-relevant knowledge, teachers also require an array of skills to enact effective practices (Baumert & Kunter, 2013; Jennings & Greenberg, 2009; Klusmann, Kunter, Trautwein, Lüdtke, & Baumert, 2008). Skill is used here as an umbrella term for social,

cognitive and affective competencies that can be learned and are thought to be important to effective teaching (Duckworth & Yeager, 2015).

1.1. Defining effective classroom teaching practices

Multiple frameworks now exist that conceptualize and allow assessment of effective classroom practices (e.g., La Paro, Pianta, & Stuhlman, 2004; Smith, Brady, & Anastasopoulos, 2008). The Classroom Assessment Scoring System (CLASS; La Paro et al., 2004) holds particular appeal. CLASS scores are predictive of long-term student achievement and more equitable teaching practices (Allen et al., 2013; Gregory et al., 2016). In addition, the CLASS factor structure generalizes across countries and student ages (Pianta & Hamre, 2009; Sandilos, DiPerna, & Investigators, 2014).

Because teaching is complex, multiple domains of enactment contribute to effective teaching. The CLASS proposes a tripartite enactment structure involving three higher-order domains: Instructional supports, Emotional supports and Classroom organization (Pianta & Hamre, 2009). Instructional supports reflect teacher behaviors intended to engender meaning-making and cognitive activation, including concept development, feedback quality, and language modeling. Emotional supports represent the overall emotional climate of the classroom as well as the quality of interpersonal relationships, indicated by

---

* Corresponding author. Center for Healthy Minds, University of Wisconsin-Madison, 625 W. Washington Avenue, Madison, WI, 53703, USA.
E-mail address: hirshberg@wisc.edu (M.J. Hirshberg).

observables such as teacher sensitivity to students and their perspectives, positive communication, and student comfort. Classroom organization refers to teacher instituted classroom structures, including interpersonal norms. Behavior management, productivity, and negative climate comprise the dimensions of Classroom organization and are assessed through indicator behaviors such as clear expectations, proactiveness, and maximizing learning time (Pianta & Hamre, 2009). Importantly, the CLASS conceptualizes effectiveness not simply by observable behaviors but primarily through the quality of interaction a behavior inspires. We theorize that teaching relevant knowledge and skill are the determinants of whether any teaching behavior becomes enactment.

1.2. Models of teacher knowledge and skills

Several conceptual models have been proposed to illustrate the knowledge and skills teachers require for enactment (Baumert & Kunter, 2013; Jennings & Greenberg, 2009; Roeser, Skinner, Beers, & Jennings, 2012). Jennings and Greenberg (2009) described the features of an effective, prosocial classroom learning environment. One of their major contributions is the explication of the teacher skills utilized during the construction of highly functional learning environments. The authors illustrate the relationship between social-emotional competencies (CASEL, 2003) such as self-awareness (e.g., identifying emotions, accurate self-perceptions), self-management (e.g., impulse control), social awareness (e.g., empathy), relationship skills (e.g., communication), and responsible decision-making (e.g., analyzing situations), enactment of effective teaching practices, teacher well-being, and salutary student outcomes. Importantly, Jennings and Greenberg's (2009) conceptualization of a prosocial classroom understands teacher social-emotional competency to be essential to both effective teaching and teacher well-being. Teacher skill is made inseparable from effective teaching practices.

Roeser et al. (2012) similarly illustrate the features of effective learning environments. These authors emphasize teacher "habits of mind" as key contributors, identifying many skills overlapping with the social-emotional competency framework. After a brief review of the relevant literature, the authors' (2012) privilege mindfulness and connection practices as particularly promising approaches to developing these skills in teachers.

Baumert and Kunter's (2013) COACTIV model is a comprehensive accounting of the inputs that support effective mathematics instruction. For example, they describe five domains of knowledge important to teaching. These five cover the spectrum of knowledge competencies that have historically been paramount within teacher education. They include content area knowledge, components of pedagogical content knowledge (e.g., knowledge of student learning processes; Shulman, 1986), and knowledge of psychology and human development. Consistent with Jennings and Greenberg (2009) and Roeser et al. (2012), Baumert and Kunter (2013) highlight the essential role of skills such as self-regulatory ability, motivational orientations, and values in effective teaching.

Knowledge and its instrumental forms in effective teaching have a long history of prominence in teacher education research (e.g., Shulman, 1986). Accordingly, teacher licensure in America and many other countries is structured around the evaluation of teaching-relevant knowledge (US Diener, Emmons, Larsen, & Griffin, 1985). Yet indicators of knowledge such as licensure scores (Goldhaber, 2007) and possession of an advanced degree show, in most cases, little to no systematic relationship to teacher performance (Rivkin et al., 2005). Further, even though teacher knowledge has been emphasized in teacher education and teacher credentialing, significant variability in teacher effectiveness persists (Chetty et al., 2014). Advances in teacher education that expand the domains of learning to include skills may offer a solution to the problem of enactment.

1.3. Mindfulness and connection practices as a paradigm for skill building to promote enactment

Mindfulness is commonly defined in the scientific literature as paying attention, on purpose, to the present moment with acceptance (Kabat-Zinn, 2013). Attention, awareness, and a kind disposition towards present moment experience (acceptance) are central mindfulness skills. Theoretical accounts of mindfulness training purport that repetitive practice builds self-regulatory capacity (e.g., attention, emotion regulation), leading to benefits including improved mental health and less emotional reactivity (Vago & Silbersweig, 2012). Connection practices such as generating feelings of warmth and goodwill for others (i.e., loving-kindness practice), wishing to relieve others' suffering (i.e., compassion practices), and contemplating the caring motivation behind an action (e.g., the decision to teach) have been shown to promote greater kindness and reduce implicit biases (Kang, Gray, & Dovidio, 2014; Weng et al., 2013).

Although nascent, research on mindfulness-based interventions with teachers indicates that integrating these practices into teacher education may support teacher well-being and enactment of effective practices. In a randomized controlled trial of teachers, increased post-test mindfulness in the training group mediated 3-month follow-up reductions in stress, depression and burnout as well occupation-related rumination while at home (Crain, Schonert-Reichl, & Roeser, 2016; Roeser et al., 2013). These data indicate that intervention-related development of the skill of mindfulness is causally related to improvement in psychological outcomes that are known to influence teaching quality (Jennings & Greenberg, 2009). However, an objective assessment of actual teaching practices was not included in this study to directly link skill development with changes in classroom teaching practices.

To date, few studies with teachers have objectively assessed the impact of mindfulness training on actual teaching practices. Jennings et al. (2017) used the CLASS (La Paro et al., 2004) in a large mindfulness RCT and found evidence for intervention group improvements on the Emotional supports and Classroom organization domains. Teachers generally score highest on Emotional supports, but theoretically it is also the domain most susceptible to stress-related diminution (Jennings & Greenberg, 2009). Consistent with this idea, a small observational study of middle school teachers found that stress and burnout were negatively correlated with mindfulness while mindfulness was positively associated with Emotional supports (Braun, Roeser, Mashburn, & Skinner, 2019). One interpretation of these data is that cultivating mindfulness may enhance Emotional supports by reducing stress, burnout and negative affectivity more generally. Jennings et al. (2017) and a small pilot conducted by Flook, Goldberg, Pinger, Bonus, and Davidson (2013) provide evidence that mindfulness training effects on teaching may extend to include organizational and instructional elements of teaching practice. Klingbeil and Renshaw (2018) concluded their meta-analysis on teacher mindfulness interventions with optimism that mindfulness training may enhance enactment of effective teaching practices.

1.4. A role for mindfulness in preservice teacher education

Mindfulness includes skills highlighted in each of the three models reviewed earlier. For example, self-regulation is core to mindfulness and the COACTIV model. In order to be mindful, one has to exercise regulation of attention and emotion. All social-emotional competency domains comprise elements of mindfulness skills. In particular, self-awareness, self-management, and social-awareness subskills such as identifying emotions, accurate self-perceptions, self-discipline, stress management, perspective-taking and empathy are focal elements of mindfulness training (Lutz, Jha, Dunne, & Saron, 2015; Vago & Silbersweig, 2012). Roeser's et al.'s (2012) model calls out mindfulness as a focal skill underlying teacher well-being and the ability to enact effective teaching practices.

Mindfulness has primarily been conceptualized as a stress and negative affect reduction training or skill. In in-service teachers, the emphasis has been placed on how mindfulness training may, by reducing the stress commonly accumulated through teaching, improve psychological functioning and thereby enhance teaching practices (e.g., Jennings et al., 2017; Roeser et al., 2013). However, undergraduate preservice teachers are presumably less stressed than professional teachers and certainly have not yet accrued the toxic levels of stress that can lead to burnout. Therefore, it is not yet known whether mindfulness will support preservice teachers in the same ways that it appears to support in-service teachers. In fact, the few studies that have examined mindfulness training during preservice teacher education find little evidence for reductions in stress or negative affect, limited evidence for increases in mindfulness, and modest evidence for increases in well-being (Garner, Bender, & Fedor, 2018; Hue & Lau, 2015). We are not aware of a mindfulness-based intervention study with preservice teachers that has assessed the impact of mindfulness training on actual teaching practices.

Consistent with the models reviewed, we predict skill development through mindfulness training will allow preservice teachers to put into practice more of what they know, filling the gap between knowledge and practice that is the problem of enactment. Teacher education and licensing regimens are structured to ensure that teachers possess adequate levels of many (though not all) forms of teaching knowledge. We contend that what is missing is systematic cultivation of the self-regulatory, attentional, empathic and other skills teachers require to take knowledge about teaching and put it into practice. Mindfulness and connection practices have shown promise as methods for developing these skills.

1.5. The present study

The primary objective of this study was to evaluate whether integrating mindfulness and connection practices into undergraduate preservice teacher education would improve enactment of effective teaching practices as assessed by the CLASS. We were secondarily interested in whether the intervention would reduce negative affectivity and enhance well-being and mindfulness skills. The multidimensional well-being (2.5.3) and negative affectivity (2.5.4) constructs were derived through factor analysis (2.6.1). The well-being construct is inclusive of Diener, Emmons, Larsen, and Griffin's (1985) notion of life satisfaction and Ryff's (Ryff, 1989) psychological well-being construct, and includes skill processes argued to be important to well-being (Kesebir, Gasiorowska, Goldman, Hirshberg, & Davidson, 2019). Negative affectivity includes negative affect as well as an array of psychological symptoms.

To advance the preservice teacher mindfulness literature, we employed a rigorous, cluster randomized controlled design that included observational, behavioral, and self-report measures. Our primary hypothesis was that mindfulness training would promote enactment of effective classroom practices as assessed by the CLASS 6-months following intervention when participants took over all teaching responsibilities during full-time student teaching. Based on the in-service teacher mindfulness intervention literature, we also hypothesized that the training would support cultivation of mindfulness skills and well-being while reducing negative affectivity.

2. Method and materials

2.1. Participants

Ninety-eight out of ~176 eligible preservice teachers (~56%) from a top-ranked elementary education preservice teacher education program (EEP) in the United States consented and pretested. One control cohort ($n = 10$) was removed prior to analysis because, in violation of study protocol, participants in this cohort took part in a yoga training that included mindfulness. The final sample was $N = 88$ ($n = 31$ Control; $n = 57$ Intervention).

The sample was representative of the demography of the teacher education program (see 2.3 for cohort information). Participants were 95.6% female ($n = 85$) and predominately Caucasian/White (86.36%, $n = 76$). About 6% of participants reported as Hispanic ($n = 5$), 3.41% reported to be of Asian ethnicity ($n = 3$) another 3.41% as African American ($n = 3$), and one participant reported as mixed or other race ($n = 1$). The average age was 21.99 years old ($SD = 0.68$). Participants had substantial prior experience working with children in some capacity ($M = 58.57$ months, $SD = 49.06$), were happy with their choice to enter into teaching ($M = 8.81$, $SD = 1.14$, out of 10), and felt that teaching was a good career fit ($M = 4.44$, $SD = .62$, out of 5).

2.2. Power analyses

A priori power analysis conducted with G*Power 3.1 software (Faul, Erdfelder, Buchner, & Lang, 2009) indicated that for planned two-tailed, post-test minus pre-test and follow-up test minus pre-test difference score models (i.e., independent group *t*-tests of GROUP × time interaction effects from mixed Analysis of Variance[ANOVA]), with α set to 0.05, power (1- β error probability) = 0.80, $N = 106$ would be required to detect moderate ($d = 0.55$) magnitude effects. Based on effect sizes from prior mindfulness research with teachers, the study was designed for an anticipated sample of $N = 100$, sufficiently powered to detect moderate and large but not small magnitude effects.

2.3. Research design

This study was a cluster RCT involving four waves (Fig. 1). The teacher education program we recruited from consists of four certification tracks: Early Childhood – English as second language; Middle childhood/Early adolescent – English as second language; Middle childhood/Early adolescent - Special education; and Middle childhood/Early adolescent – Content focused minor. For four consecutive semesters (Fall 2015–2017), we recruited from the two cohorts entering their third semester in the program. In total, two of each type of the four licensure tracks were recruited into the study (i.e., 8 cohorts total) (see Fig. 2).

Because we were interested in the effects of integrating mindfulness into teacher education rather than mindfulness training separate from it, the intervention was partially delivered during weekly cohort seminar, ruling out random assignment of individuals. We instead randomized by cohort. To control for the possibility that licensure tracks enrolled students with different characteristics or systematically influenced students differently, we block randomized by licensure track via a random number generator so that at study end one cohort from each licensure track was assigned to intervention and the other to control. Participants continue to be followed through 2020 to examine the impact of preservice mindfulness training on the transition to professional teaching.

Pretesting occurred at the beginning of the first semester in the study, prior to intervention onset. Baseline teaching observations were collected during practicum (i.e., part-time student teaching), but because practicum begins a few weeks into the semester, about 30% of baseline observations occurred after the first intervention session had occurred. Post-testing was completed in a 2-week period immediately after intervention completion (no CLASS at post-test). Follow-up testing occurred at the end of the subsequent semester (i.e., final undergraduate semester), between 5 to 7-months after post-testing. The final CLASS observation was conducted during 'lead weeks.' Lead weeks is the 2 to 4-week period in which participants took over all teaching responsibilities from their supervising teacher.
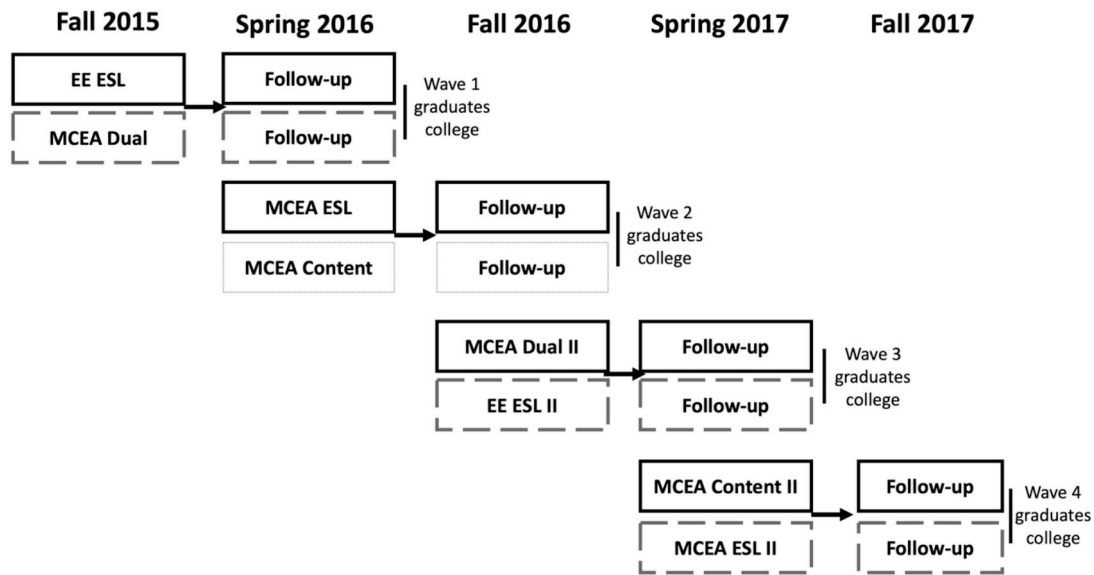
Note: Black outline = intervention cohort. Dashed outline = control cohort. Gray outline = control cohort removed for receiving yoga/mindfulness training. EE ESL = Early Education English as Second Language certification cohort; MCEA = Middle Childhood / Early Adolescent; Dual = dual certification cohort; ESL = English as Second Language certification; Content = Content area certification.

Fig. 1. Study design.

Note: Black outline = intervention cohort. Dashed outline = control cohort. Gray outline = control cohort removed for receiving yoga/mindfulness training. EE ESL = Early Education English as Second Language certification cohort; MCEA = Middle Childhood/Early Adolescent; Dual = dual certification cohort; ESL = English as Second Language certification; Content = Content area certification.

2.4. Procedure

All study procedures were approved by the university institutional review board. The study design, measures, and analysis approach were preregistered (clinicaltrials.gov/ct2/show/NCT02544412). Participants were recruited through emails and an in-class visit. At pretest, participants were met by trained undergraduate research assistants who provided a printed copy and read the consent form. All participants provided written, informed consent before the study began and were compensated $145 for completing the study.

A Latin square design was used to construct six testing orders. Each computerized behavioral task represented a single block; the self-reports collectively represented another block. Self-reports were completed last, with demographic questions at the end. The order of behavioral tasks was counterbalanced across groups. Behavioral task blocks were comprised of: Emotional Go/No-go (see 2.5.2); Breath Count task (Levinson, Stoll, Kindy, Merry, & Davidson, 2014); and a block comprised of the Adult and Child versions of the Black/White Implicit Association Test (Lane, Banaji, Nosek, & Greenwald, 2007). Breath count task results were not analyzed because subsequent research has called into question the task's construct validity as well as scoring methods (Hirshberg et al., 2018). Results from the IATs are reported separately (Hirshberg, Flook, Enright, & Davidson, unpublished manuscript).

Testing was conducted on Dell Latitude E5550 15" monitor laptops using EPrime version 2.1.0 (Psychological Software Tools, 2012) and Qualtrics (Qualtrics, 2015). Participants always followed the same testing order. All study procedures were consistent across waves. We maintained a single-blind design by keeping research assistants and CLASS coders blind to group assignment and study hypotheses.

2.4.1. Well-being skills training

The 9-week training consisted of weekly 1.5-h classes along with

two 4-h intensive days (21.5-h total). Thirty minutes of the intervention was integrated into the mandatory weekly cohort seminar. The remaining hour of training occurred either before or after seminar, depending on cohort scheduling. Participants were asked to meditate for 10–20 min/day. Over follow-up, intervention cohorts engaged in weekly 15-min practices via audio recording with an instructor present to lead once per month. The same three instructors, all former classroom teachers (> 40 years combined experience) with extensive mindfulness teaching (> 5 years each) and practice (> 10 years each) experience, taught all intervention waves.

The training has roots in Mindfulness-based Stress Reduction (MBSR; Kabat-Zinn, 1982) and Tibetan Buddhist approaches to mind-training. We added to common secular presentations of mindfulness a novel element – a secularized theory of change derived from Buddhist philosophy (Thrangu & Thrangu, 2004). In Buddhist philosophy, a schema of the process of training the mind is considered an important catalyst for accruing benefits from contemplative practice (Rinpoche, 1993). In the Tibetan meditation tradition, this is known as proper view and comprises one of the three major tenets of mind-training (the others are meditation and conduct/ethics). The secular view in this training was intended to provide new mindsets for approaching mindfulness and connection practices. View components were provided as short (~5 min) didactic content, reinforced during meditation instruction.

As an example, the view in the third class was that desired qualities of mind like the ability to be at rest even with difficult experiences (i.e., equanimity) are already present but habitual approaches to experience (e.g., aversion to negative) prevent them from being noticed. The intent of this view is to shift intention away from changing experience toward an openness to experiencing whatever is arising as it is. Instructors went on to state that in allowing experience to be what it is, a quality of calmness naturally arises even during challenge (see supplementary materials for overview and sample lesson). In contrast to many secular MBIs that highlight stress reduction, the present training emphasized
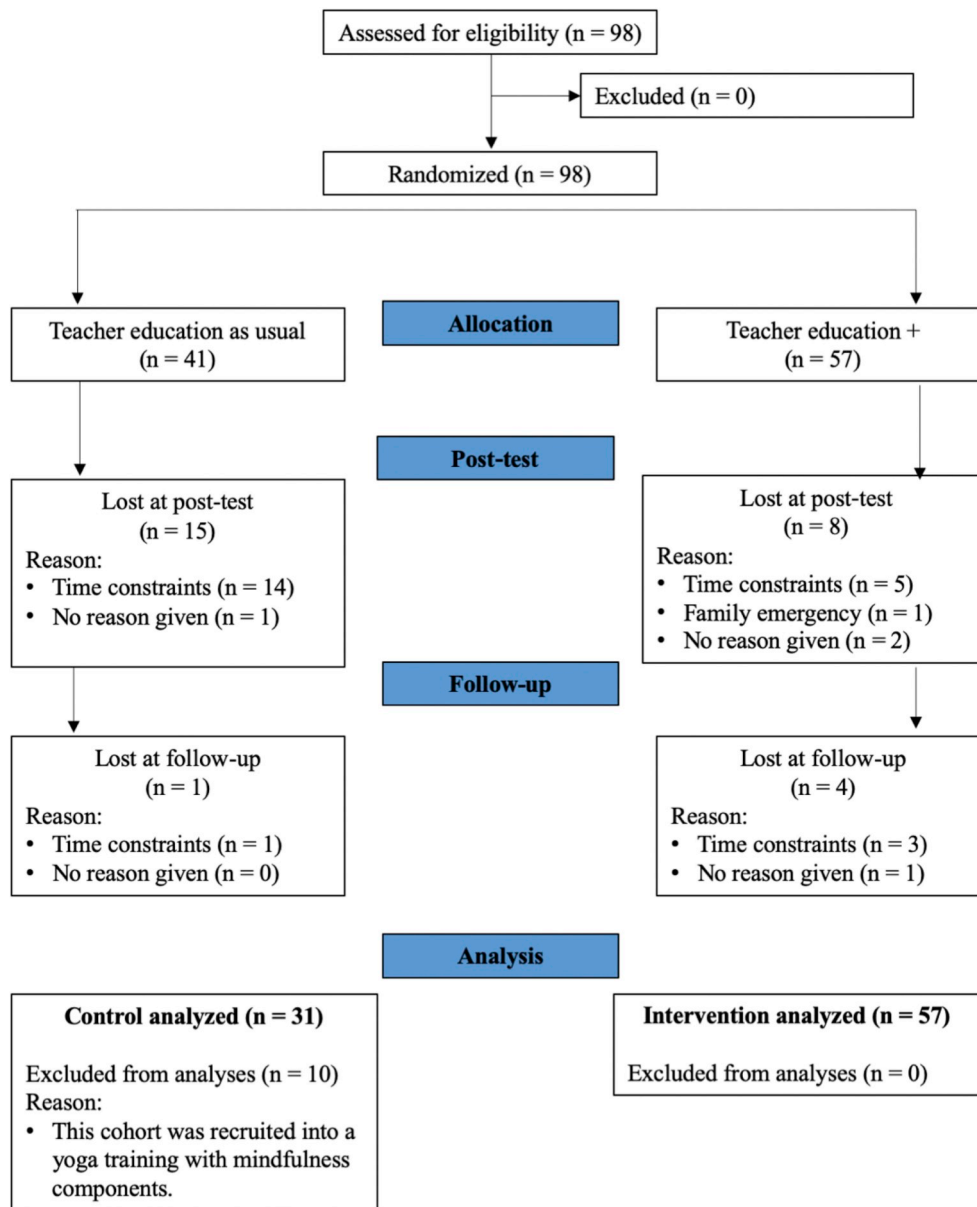
Fig. 2. CONSORT diagram.

instead that in allowing experience to be what it is, the content of experience becomes less stressful.

Consistent with other teacher mindfulness-based interventions (e.g., CARE; Jennings, Snowberg, Coccia, & Greenberg, 2011), the present intervention utilized mindfulness, loving-kindness and compassion practices. Mindfulness practices included several forms of meditation intended to develop focused and flexible attention imbued with an attitude of acceptance. Loving-kindness practices involved reflection on others' desire to be happy and safe and generating the wish that they are able to experience happiness and safety. Compassion practices involved reflecting on the suffering others experience and generating a desire to relieve others of their suffering (Dahl, Lutz, & Davidson, 2015; Weng et al., 2013). The present intervention also contained a series of contemplations on the caring motivation that led participants to enter the preservice program and strategies for strengthening that caring motivation.

2.5. Measures

Measures were selected to assess: a) actual classroom teaching practices; b) mindfulness skills; c) well-being skills theoretically important to enactment; and d) negative affect symptoms commonly elevated among in-service teachers. Negative affect variables and some well-being measures were primarily intended to model growth trajectories longitudinally when data collection is complete (i.e., in 2020). However, our pre-registration did not clearly state this point. To avoid potential biases associated with selective reporting of measures, we report on all measures including those intended to be used in longitudinal analyses. To reduce the number of inferential tests, we subjected all measures except mindfulness (because of the primacy of this outcome in a mindfulness intervention study) to a factor analysis (see Jennings et al., 2017 and 2.6.1 for details) that resulted in three latent factors (the CLASS, negative affectivity, and multidimensional well-being).

2.5.1. Objective ratings of classroom teaching practices

The Classroom Assessment Scoring System (CLASS; La Paro et al., 2004) is a standardized, objective observational tool that quantifies the quality of classroom teaching practices and interactions (Allen et al., 2013). Between the four EEP cohorts, state licensure is provided in grades pKindergarten - Eight, translating into four levels of the CLASS: Pre-K, K-3, Upper elementary (i.e., grades 4–6), and Secondary (i.e., grades 7–12). As noted in 1.4, CLASS's three higher-order domains generalize across levels. We report only on these domains. Domain scores range from 1 to 7. Higher scores represent greater domain mastery. Observations were conducted according to CLASS recommendations by certified coders who were blind to group assignment. Due to scheduling limitations, baseline data were collected on 80/88 (91%) participants.

2.5.2. Mindfulness

The Five Facet Mindfulness Questionnaire (FFMQ; Baer et al., 2008) is a 39-item questionnaire measuring five mindfulness competencies. Observe ($\alpha = .77$) assesses awareness of cognitions, emotions, and sensory aspects of experience. Describe ($\alpha = .89$) assesses the ability to articulate experience. Act with Awareness ($\alpha = .91$) assesses the degree to which actions are done with awareness and intention. Non-judging of inner experience ($\alpha = .92$) measures the degree to which cognitions, emotions and sensory experiences are judged critically (all items reverse-scored). Non-react to inner experience ($\alpha = .77$) assesses the ability to be aware of but not reactive to inner experience. The FFMQ uses a Likert like scale ranging from 1 (Never or very rarely true) to 5 (Very often or always true). Facet scores range from 8 to 45. Higher scores reflect greater mindfulness.

2.5.3. Multidimensional well-being factor

The Emotional Go/No-go (Hare et al., 2008) assesses behavioral regulation over emotional cues. Participants are presented with a series of happy, neutral, or fearful adult faces. On go trials, participants hit the j key as quickly as possible, inhibiting the key press on nogo trials. All possible combinations of responses (i.e., go, nogo) and stimuli (happy, neutral, fearful) are used (i.e., six total blocks). Faces appear for 500 ms followed by a randomly alternating inter-stimulus interval (ISI) ranging between 500 and 1500 ms (mean 750 ms) during which participants rest viewing a fixation cross. Dprime, an index of accuracy accounting for response bias, is the outcome score.

The current version of the Emotional Style Questionnaire (ESQ: Kesebir et al., 2019) is a 48-item scale containing six subscales that are constructed as well-being competencies: Resilience (i.e., rapidity of recovery from a challenge; $\alpha = .74$), Positive outlook (i.e., disposition to view experiences positively or with meaning; $\alpha = .71$), Social intuition (sensitivity to social cues; $\alpha = .82$), Self-awareness (i.e., awareness of internal experience and reactions; $\alpha = .60$), Sensitivity to context (adaptation to social context; $\alpha = .64$), and Attention (i.e., ability to direct and sustain attention; $\alpha = .80$). Only the total score ($\alpha = .87$), representing overall healthy emotionality, was analyzed. Participants rate items on a Likert-like scale ranging from 1 (strongly disagree) to 7 (strongly agree). Higher scores reflect greater healthy emotionality.

The General Self-efficacy Scale (Jerusalem & Schwarzer, 1992) is a 10-item scale that assesses beliefs about one's ability to manage life's demands ($\alpha = .81$). Scores range from 10 to 40, with higher scores reflecting greater self-efficacy. On a 1 (not at all) to 4 (exactly true) scale, participants answer items including "I can usually handle whatever comes my way" and "It is easy for me to stick to my aims and accomplish my goals." General self-efficacy rather than teaching specific self-efficacy was selected because we were interested in MBI impacts on global self-perception and the two scales are highly correlated in teachers (Lauermann & König, 2016).

The Implicit Positive and Negative Affect Test (Quirin, Kazén, & Kuhl, 2009) positive subscale measures implicit positive affect by asking participants to rate the extent to which artificial words express positive moods. On a 4-choice scale ranging from "doesn't fit at all" to "fits very well", participants rate artificial words based on three positive (happy, cheerful, energetic) adjectives. Implicit positive affect is the average of endorsed positive ($\alpha = .79$) adjectives. Higher scores represent more implicit affect.

The Maslach Burnout Inventory – Education Survey (Maslach, Jackson, & Leiter, 1996) personal accomplishment subscale ($\alpha = .53$) uses 8-items to assess the degree to which one is satisfies with their occupational performance. Scores range from 0 to 48, respectively with higher scores representing greater personal accomplishment.

The Positive and Negative Affect Schedule (Watson, Clark, & Tellegen, 1988) positive subscale is a 10-item measure of general positive affect. On a 1 (very slightly/not at all) to 5 (extremely) scale, participants rate how much positive (e.g., interested, proud) feeling they generally experience. Positive affect ($\alpha = .87$) is the sum scores of respective items. Higher scores reflect greater positive affect.

The Psychological Well-Being scale (Ryff & Keyes, 1995) is a 42-item multi-dimensional assessment of psychological well-being. Only the total score was used ($\alpha = .94$). Scores range from 42 to 252 with higher scores reflecting greater overall well-being.

2.5.4. Negative affectivity factor

The Implicit Positive and Negative Affect Test (Quirin et al., 2009) negative subscale measures implicit negative affect by asking participants to rate the extent to which artificial words express negative adjectives (helpless, tense, inhibited). Implicit negative affect ($\alpha = .63$) is the average of endorsed negative adjectives. Higher scores reflect more implicit negative affect.

The Positive and Negative Affect Schedule negative subscale (Watson et al., 1988) is a 10-item measure of general negative affect. On a 1 (very slightly/not at all) to 5 (extremely) scale, participants rate how much negative (e.g., guilty, afraid) feeling they generally experience. Negative affect ($\alpha = .81$) is the sum scores of respective items. Higher scores reflect greater negative affect.

The Maslach Burnout Inventory – Education Survey (Maslach et al., 1996) Emotional exhaustion ($\alpha = .87$) and Depersonalization ($\alpha = .80$) subscales assess the emotional fatigue and lack of interpersonal connection indicative of occupational burnout. Scores range from 0 to 54 to 0–30 on these subscales, respectively. Higher scores represent greater endorsement of the subscale construct.

The Perceived Stress Scale (Cohen, Kamarck, & Mermelstein, 1983) is a 10-item self-report assessing feelings of stress over the prior month ($\alpha = .83$). It employs a 5-point Likert-like scale ranging from 0 (never) to 4 (very often). Scores range from 0 to 40, with higher scores representing higher levels of perceived stress.

The Symptoms Checklist 90-R (Derogatis, 1992) is a 90-item self-report inventory used to measure psychological distress. The scale assesses nine symptom areas and three symptomatology aggregates. Symptomatology aggregates combine multiple symptom areas to assess a general area of symptoms (E.g., the global severity of all symptoms). Only the anxiety ($\alpha = .80$) and depression ($\alpha = .86$) subscales were used because these are commonly elevated in in-service teachers. Scores are the average of scale items (range = 0–4); higher scores represent greater endorsement of the symptom area.

2.5.5. Daily mindfulness practice reports (intervention only)

Intervention participants completed a daily log of mindfulness practice each week through an online database. Forty-five out of 57 (78.94%) intervention participants turned in at least one week of the practice log. Over 91% of the intervention participants that completed post-testing provided one week or more of practice logs. Three participant reports were excluded from analyses because reported practice was > 4 SD above the mean. On average, participants reported on 35.55 days of practice over the intervention period.

2.6. Analytical strategy

Primary analyses consisted of independent group Welch's *t*-tests on post-test minus pre-test and follow-up test minus pre-test difference scores. Difference score *t*-tests are equivalent to mixed ANOVA inter-action effects (i.e., GROUP x time) except that Welch's *t*-tests are used in place of Student's *t*-test. Welch's *t*-tests are suggested because they do not assume equal samples sizes or variances (Delacre, Lakens, & Leys, 2017). We report Cohen's *d* with 95% confidence intervals as an esti-mate of effect size and its variability.

2.6.1. Identifying latent factors through factor analysis

We reduced dimensionality following a procedure described in Jennings et al. (2017). Using baseline data, we first examined inter-correlations within the theory driven constructs of multidimensional well-being, negative affectivity, and enactment (Supplementary mate-rials Table 1). Exploratory factor analysis (EFA) with promax oblique rotation and principal axis factoring resulted in a set of three factors coherent with our *a priori* theory (Supplementary materials Table 2; Fabrigar, Wegener, MacCallum, & Strahan, 1999). The first factor that explained the majority of variance corresponded to negative affectivity. It consisted of explicit negative affect (PANAS negative subscale), emotional exhaustion and depersonalization (MBI subscales), perceived stress, and depressive and anxiety symptoms (SCL-90 R subscales). The second factor consisted of the three domains of the CLASS. Even though we intended to analyze the CLASS according to its three major domains (2.5.1), we included them in the factor analysis to determine whether our theoretical model was justified in these data. The final factor cor-responded to our multidimensional well-being construct and consisted of explicit and implicit positive affect (PANAS positive subscale), self-efficacy, personal accomplishment (MBI subscale), psychological well-being (PWB total score), and healthy emotionality (ESQ total score). Neither Dprime from the Emotional Go/No-go nor the IPANAT sub-scales loaded adequately on a factor. The IPANAT also exhibited an unexpected positive correlation between the positive and negative subscales and low internal consistency. Both the IPANAT and the Emotional Go/Nogo were removed from analyses. EFA factor scores were used for all analyses.

We next conducted confirmatory factor analysis (CFA) on each factor to determine adequate fit to our theoretically and empirically derived constructs. Fit indices (e.g., TLI, CFI, RMSEA, SRMR) were all in the good to adequate range (Supplementary materials Table 3). We limited inferential tests to the CLASS domains, FFMQ facets, negative affectivity, and multidimensional well-being.

2.6.2. Missingness

Missing data is common but problematic in RCTs because the missingness approach can affect inferential statistics (Graham, 2009; White, Horton, Carpenter, Pocock, & others, 2011). Multiple imputation has been shown: a) in simulation studies to maintain type I error rates; b) be effective in small samples even when data are not Missing at Random; and c) to perform well with substantially higher rates of missingness than the present data (Graham, 2009; Graham & Schafer, 1999; Rubin, 2004; Sinharay, Stern, & Russell, 2001). In the present study, post-test missingness was about 35% ($n = 12$) in the control and 14% ($n = 8$) in the intervention groups. At the 6-month follow-up, control and intervention missingness was around 40% ($n = 13$) and 20% ($n = 12$), respectively.

We conducted multiple imputation via Multivariate Imputation with Chained Equations, imputing 50 datasets via predictive mean matching with the seed set to 1981 for replicability (Buuren & Groothuis-Oudshoorn, 2011). Subsequent analyses followed Rubin's rule for pooling repeated analyses (Rubin, 2004). Sensitivity analysis was con-ducted comparing the planned Intent-to-treat (ITT) analysis with mul-tiple imputation to per protocol analysis (i.e., List wise deletion). Effect sizes were equivalent between approaches (Supplementary materials

Table 4). Only the ITT results are presented. All data processing and analysis was performed in R (Team, 2014). All data and the syntax used to conduct reported analyses are accessible online (osf.io/5pjnh/).

3. Results

Independent group *t*-tests on baseline data revealed no significant between-group differences ($p > .05$) on any variable.

3.1. Intervention fidelity

Around 70% (39/57) of treatment participants reached criteria for full class attendance (i.e., 8/9 classes attended) and ~90% (50/57) attended at least one intensive day. We examined content im-plementation fidelity by providing instructors with checklists of the core content intended to be covered after each class. Over 95% of content was delivered during the intended session, and there was near total consistency (> 95%) between the waves. All but one cohort, due to external disruptions, followed the prescribed 15-min follow-up practice schedule.

3.2. Classroom management behaviors

At the 6-month follow-up coinciding with full-time student teaching, difference scores (follow-up minus pre-test) on all CLASS domains were significantly different between groups, with moderate magnitude effects favoring the intervention on Instructional supports *t* (45.15) = 2.98, $p = .005$ $d = 0.73$ CI[0.23, 1.23], Emotional supports *t* (29.64) = 2.24, $p = .033$, $d = 0.55$ CI[0.05, 1.04], and Classroom organization *t*(41.78) = 2.20, $p = .033$, $d = 0.54$ CI[0.04, 1.03].

3.2.1. Daily mindfulness practice is associated with enactment gains

Participants reported practicing mindfulness during the interven-tion for an average of 13.94 min each day ($SD = 9.47$). Daily reported mindfulness practice was significantly associated with improved CLASS scores on Instructional supports $r = .39$, $p = .007$ and Classroom or-ganization $r = .38$, $p = .011$. Associations with Emotional supports were also moderate in magnitude but did not reach statistical sig-nificance, $r = .25$, $p = .078$ (Fig. 3). Daily reported mindfulness practice explained 14%, 13%, and 7% of intervention group improve-ments on the CLASS domains, respectively (see Fig. 4).

3.3. Mindfulness

The intervention group showed a significantly greater increase from pre to post-test on the Observe subscale of the FFMQ *t*(45.98) = 2.18, $p = .035$, $d = 0.49$ 95% CI[0.4, 0.94]. A similar magnitude interven-tion group increases was evident on the Nonreact subscale from pre-to post-test *t*(45.66) = 1.98, $p = .054$, $d = 0.44$ CI[-0.01, 0.89]. A small magnitude, non-significant intervention group increases was present on the Describe subscale from pre to post-test *t*(41.99) = 1.50, $p = .140$, $d = 0.33$ CI[-0.11, 0.78]. No differences were evidenced on the Act with awareness or Nonjudge subscales *t*(47.08) = −0.23, $p = .817$, $d = −0.05$ CI[-0.50, 0.39] and *t*(43.06) = −0.69, $p = .496$, $d = −0.15$ CI[-0.60, 0.29], respectively. No follow-up minus pre-test group differences were observed on any FFMQ subscale (all *p*s > .10).

3.4. Multidimensional well-being and negative affectivity

We observed no post-test minus pre-test between-group differences in well-being *t*(26.87) = −0.17, $p = .865$, d −0.04 CI[-0.48, 0.41] or negativity affectivity *t*(38.10) = 0.31, $p = .757$, $d = 0.07$ CI[-0.37, 0.52]. Similarly, follow-up test minus pre-test group differences were not observed on well-being or negative affectivity at follow-up, *t* (38.59) = 0.71, $p = .483$, $d = −0.16$ CI[-0.6, 0.29] and *t* (35.50) = 0.37, $p = .710$, $d = 0.08$ CI[-0.36, 0.53].
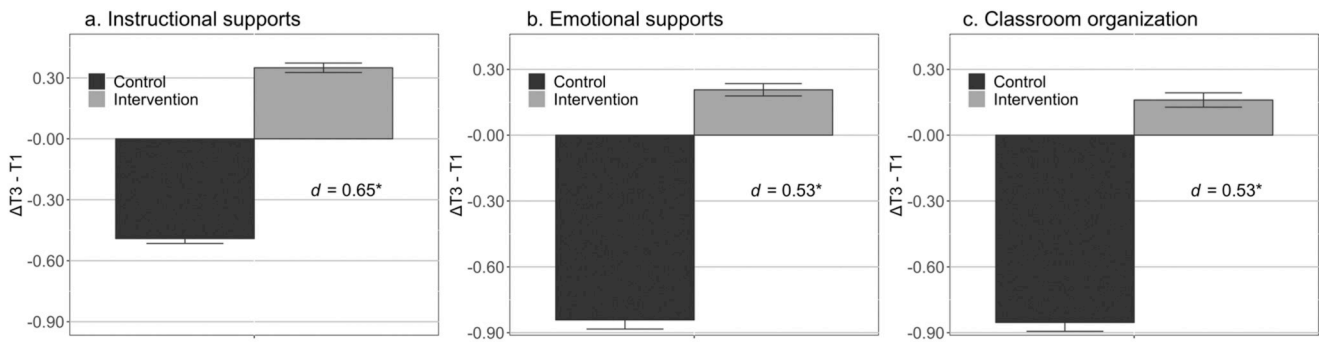
Note: a. GROUP x time interaction on CLASS Instructional supports. b. GROUP x time interaction on CLASS Emotional supports. c. GROUP x time interaction on CLASS Classroom organization. Error bars = SEM. d = Cohen's d effect size. *** $p < .001$, ** $p < .01$, * $p < .05$.

Fig. 3. Six-month follow-up minus pre-test changes on the CLASS.
Note: a. GROUP × time interaction on CLASS Instructional supports. b. GROUP × time interaction on CLASS Emotional supports. c. GROUP × time interaction on CLASS Classroom organization. Error bars = SEM. d = Cohen's d effect size. ***$p < .001$, **$p < .01$, *$p < .05$.

4. Discussion

At the 6-month follow-up timed to coincide with preservice teacher participants taking over classroom responsibilities during full-time student teaching, between-group change was significantly different and favored the intervention group on all major domains of the CLASS: Instructional supports, Emotional supports, and Classroom organization. CLASS scores have been associated with salutary teacher and student outcomes and are longitudinally predictive of student achievement (e.g., Allen et al., 2013). These data provide the first evidence that training preservice teachers in mindfulness and connection practices supports preservice teacher enactment of effective teaching across grade levels from K - 8. Furthermore, consistent with research on skill acquisition (Chi, Glaser, & Farr, 2014), daily reported mindfulness practice in the intervention group was significantly associated with improvements on Instructional supports and Classroom organization, and moderately but not significantly associated with improvements on Emotional supports. Combined, these results indicate that greater engagement in mindfulness practice is associated with more effective classroom teaching practices among preservice teachers learning to teach.

On all of the CLASS domains except Instructional supports, intervention group improvements were smaller than control group decrements. Lower CLASS scores at follow-up were expected because baseline observations were collected during practicum. Participants supported instruction during practicum (and were the focus of the observation) but were not always directing the classroom. In contrast, the second observation was conducted when participants had taken over all classroom responsibilities. One would expect follow-up scores may be

lower because they reflect participant teaching behaviors only. It is notable then that the intervention group improved at follow-up whereas the control group showed decrements on all CLASS domains. Further, intervention group follow-up scores were closer to experienced in-service teachers than to controls (US Department of Health and Human Services, 2019; Teachstone, 2011).

This study has several strengths. Block randomizing by cohort type resulted in the four cohort types serving as their own control, controlling for several important confounds. Because the content of the licensure tracks is standardized and dictated partly by exogenous factors (e.g., state licensure requirements), participants in the same licensure tracks were exposed to identical teacher education content knowledge. A second strength of the study is the pre-, post- and 6-month follow-up design. The lack of follow-up assessments in most teacher and all preservice teacher mindfulness research has left a gap in our understanding of the persistence and/or emergence of effects in the period following intervention. These data indicate that long-term behavioral changes (i.e., the CLASS) may be observed absent concomitant self-report changes.

There are several potential explanations for the discrepancy between the CLASS and self-report findings. One possibility is that unobserved group differences in characteristics related to teaching practices but not affected by mindfulness training drove intervention group improvements on the CLASS. At the suggestion of a reviewer, we examined whether Big Five personality characteristics (collected as a baseline exploratory measure; Rammstedt & John, 2007) were different between groups or related to gains on the CLASS. Personality has been found to predict persistence in teaching and correlates at small magnitudes with various metrics of teacher performance (Jamil, Downer, &
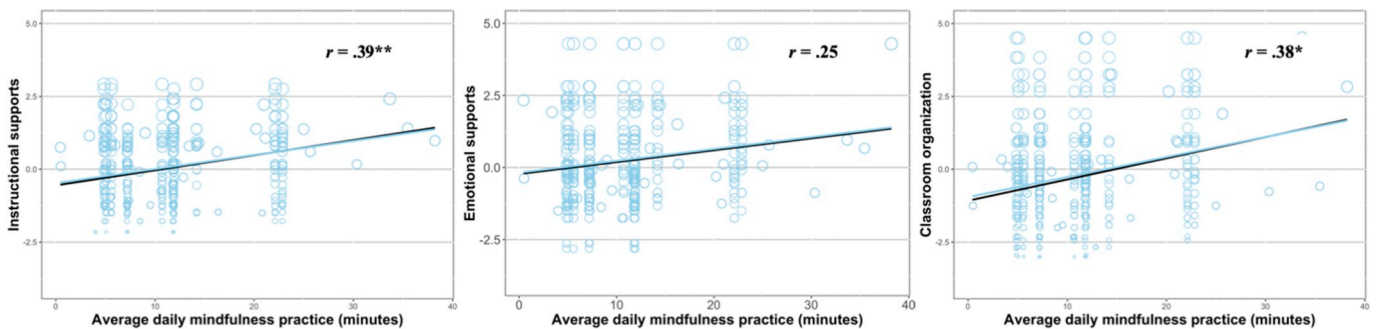

Fig. 4. Daily reported mindfulness practice is associated with improvements on the CLASS.
Notes: Black line = slope for observed data. Blue line = slope for multiply imputed data. Size of circle reflects likelihood of imputed value, with large circles more commonly imputed. ***$p < .001$, **$p < .01$, *$p < .05$.

Pianta, 2012; Kim, Jörg, & Klassen, 2019). The groups in this study did not differ on any personality factor at baseline. Although agreeableness was significantly associated with improvements on Instructional supports ($r = .24$) and Classroom organization ($r = .23$), the interaction between agreeableness and GROUP was not significant on either domain (Supplementary materials Table 6). Although it remains possible that unobserved variables affected the intervention group differentially, many teacher characteristics known to be related to teacher performance (e.g., self-efficacy, personality) were assessed. Coupled with the block randomized study design and the lack of group differences at baseline on observed variables, this is an unlikely explanation for the CLASS results.

4.1. Potential differences in mindfulness effects based on professional status

The prevailing hypothesis undergirding most teacher mindfulness research is that mindfulness training reduces teacher stress and negative affect thereby improving teaching practices. This hypothesis may be valid in in-service teachers experiencing heightened stress but be invalid for preservice teachers who have not yet been exposed to the full challenges of the profession. In this sample, baseline means on perceived stress, negative affect and psychological symptoms were below the low-to-moderate range based on scale breakpoints. Prior mindfulness research with preservice teachers has reported null findings on negative affectivity (Hue & Lau, 2015). Divergence in pre- and in-service teacher outcomes may indicate mindfulness operates on different mechanisms based on professional status. Understanding more about the pathways by which mindfulness interventions impact teachers and whether these differ based on professional status is an important area for future research.

This research was designed around an implied mediation model whereby increases in skill promote enactment of effective teaching practices. Although the present study was not powered to detect mediation effects nor were these effects explicitly hypothesized, we do not find support for such a model in these data. In addition to observing no intervention related decreases in negative affectivity, post-test gains in mindfulness were no longer evidenced at follow-up.

When interpreting these results, it is important to note that the lack of follow-up intervention group gains in mindfulness relative to controls was primarily the result of increased self-reported mindfulness in the control group. That is, the intervention group did not experience a decline in self-reported mindfulness from post-to follow-up. Rather, for unknown reasons, the control group reported an increase. Measurement invariance between meditating and non-meditating samples has been observed on self-reports of mindfulness, including the Five Facet Mindfulness Questionnaire used here (Van Dam, Earleywine, & Danoff-Burg, 2009). Thus, the unexpected increase in control group mindfulness at follow-up may reflect participant bias or measurement error associated with self-reports of mindfulness. A related possibility is that mindfulness training resulted in response shift bias (Howard & Dailey, 1979), altering intervention participants' frame of reference such that endorsement of mindfulness items was tempered at follow-up. Some research has found that mindfulness self-reports fail to differentiate between mindfulness and other interventions (Goldberg et al., 2015). A less likely possibility is that teacher education itself is an intervention that increases mindfulness facets. This discussion parallels an area of recent interest in the field; the paucity of valid and reliable measures to assess the skills cultivated through mindfulness training (Van Dam et al., 2017).

In the case that mindfulness interventions operate on different pathways in pre- and in-service teachers, it is not surprising that the intervention group did not decrease in negative affect at the same time that improvements on the CLASS were evidenced. Beginning to teach is exhausting and difficult. The post-test intervention group improvements in facets of mindfulness may have provided the foundation to apply greater attention, awareness, and effort during student teaching,

resulting in higher CLASS scores. But human beings possess limited resources to expend on learning (Feldon, 2007). The cost of this attention, awareness and effort during the initial phase of learning to teach might understandably manifest in *higher* stress (McCann & Johannessen, 2004). We expect that this pattern would invert after some period of adaptation and learning. As new teachers become more competent in the basic practices associated with effective teaching, these practices become more automated, requiring less effort and freeing up resources for more advanced elements of instruction (Baartman & De Bruijn, 2011; Baumert & Kunter, 2013; Berliner, 2001; Feldon, 2007) Over the long-term the initial costs associated with improved enactment would be expected to be offset by longer-term benefits derived from establishing a functional classroom environment.

4.2. Limitations

Several limitations are important to note. An active control condition would provide a more rigorous comparison of intervention effects and help rule out the so-called common factors explanation (Wampold, 2015). Randomizing at the participant level or powering the study to estimate clustering effects through hierarchal linear modeling is preferred with clustered data but was not feasible in the present study. Future research with very large samples would allow adequate power to detect potentially small magnitude but meaningful differences in group change, mediation effects and afford estimation through hierarchal linear modeling.

Although we employed measures at several levels of analysis (i.e., self-report, cognitive behavioral, objective observational), there are few valid and reliable measures that assess with specificity mindfulness-related skills (Van Dam et al., 2017). This leaves open the possibility that some effects were unobserved because of inadequate measurement. Similarly, we were not able to compute CLASS coder inter-rater reliability. Our large pool of certified coders was blind to participant condition and each coder rated a small number of participants, making systematic coder bias unlikely, but we cannot eliminate this possibility. Finally, despite efforts to retain participants (e.g., email, phone call), attrition was higher than desirable, particularly in the control group.

4.3. Conclusions

In an experimental design that ensured participants were exposed to equivalent teacher education knowledge, this study provides the first evidence that mindfulness training during preservice teacher education leads to significant improvements in effective classroom teaching practices. The mechanisms undergirding the observed improvements in teaching practices remain to be explicated as does the impact of mindfulness training on the transition to professional teaching. Insofar as difficulty with classroom management is a major cause of stress and burnout among new teachers, the finding that mindfulness training supports preservice teachers' teaching practices, including practices related to classroom management, suggests that mindfulness training may be one approach to making the transition to professional teaching less stressful by affording higher levels of instructional skill.

Author note

Matthew J. Hirshberg, Center for Healthy Minds and Department of Educational Psychology, University of Wisconsin Madison; Lisa Flook, Center for Healthy Minds, University of Wisconsin Madison, Robert D. Enright, Department of Educational Psychology, University of Wisconsin Madison, Richard J. Davidson, Center for Healthy Minds, University of Wisconsin Madison.

Lisa Flook is now at the Learning Policy Institute, Palo Alto, California.

RDE is a Board Member, International Forgiveness Institute, Inc., Madison, WI. RJD is the founder, president, and serves on the board of

directors for the non-profit organization, Healthy Minds Innovations, Inc. In addition, RJD served on the board of directors for the Mind & Life Institute from 1992 to 2017. No donors, either anonymous or identified, have participated in the design, conduct, or reporting of research results in this manuscript.

Acknowledgment

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2019.101298.

References

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observations of effective teacher-student interactions in secondary school classrooms: Predicting student achievement with the classroom assessment scoring system-secondary. *School Psychology Review, 42*(1), 76.

Baartman, L. K., & De Bruijn, E. (2011). Integrating knowledge, skills and attitudes: Conceptualising learning processes towards vocational competence. *Educational Research Review, 6*(2), 125–134. https://doi.org/10.1016/j.edurev.2011.03.001.

Baer, R. A., Smith, G. T., Lykins, E., Button, D., Krietemeyer, J., Sauer, S., et al. (2008). Construct validity of the five facet mindfulness questionnaire in meditating and nonmeditating samples. *Assessment, 15*(3), 329–342. https://doi.org/10.1177/1073191107313003.

Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 25–48). Springer. https://doi.org/10.1007/978-1-4614-5149-5_2.

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research, 35*(5), 463–482. https://doi.org/10.1016/S0883-0355(02)00004-6.

Braun, S. S., Roeser, R. W., Mashburn, A. J., & Skinner, E. (2019). Middle school teachers' mindfulness, occupational health and well-being, and the quality of teacher-student interactions. *Mindfulness, 10*(2), 245–255. https://doi.org/10.1007/s12671-018-0968-2.

Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), Retrieved from http://doc.utwente.nl/78938/.

Cheng, A., & Zamarro, G. (2018). Measuring teacher non-cognitive skills and its impact on students: Insight from the measures of effective teaching longitudinal database. *Economics of Education Review, 64*, 251–260. https://doi.org/10.1016/j.econedurev.2018.03.001.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *The American Economic Review, 104*(9), 2633–2679. https://doi.org/10.1257/aer.104.9.2633.

Chi, M. T., Glaser, R., & Farr, M. J. (2014). *The nature of expertise.* Psychology Press.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*(4), 385–396. https://doi.org/10.2307/2136404.

Crain, T. L., Schonert-Reichl, K. A., & Roeser, R. W. (2016). Cultivating teacher mindfulness: Effects of a randomized controlled trial on work, home, and sleep outcomes. *Journal of Occupational Health Psychology*. https://doi.org/10.1037/ocp0000043.

Dahl, C. J., Lutz, A., & Davidson, R. J. (2015). Reconstructing and deconstructing the self: Cognitive mechanisms in meditation practice. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2015.07.001.

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of Student's t-test. *International Review of Social Psychology, 30*(1).

Derogatis, L. R. (1992). SCL-90-R: Administration, scoring and procedures manual for the R (evised) version and other instruments of the psychopathology rating scale series. *Clinical Psychometric Research,* 1–16.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*(4), 237–251. https://doi.org/10.3102/0013189X15584327.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272. https://doi.org/10.1037//1082-989x.4.3.272.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149.

Feldon, D. F. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist, 42*(3), 123–137. https://doi.org/10.1080/00461520701416173.

Flook, L., Goldberg, S. B., Pinger, L., Bonus, K., & Davidson, R. J. (2013). Mindfulness for teachers: A pilot study to assess effects on stress, burnout and teaching efficacy. *Mind, Brain and Education: The Official Journal of the International Mind, Brain, and Education Society, 7*(3), https://doi.org/10.1111/mbe.12026.

Garner, P. W., Bender, S. L., & Fedor, M. (2018). Mindfulness-based SEL programming to increase preservice teachers' mindfulness and emotional competence. *Psychology in the Schools, 55*(4), 377–390. https://doi.org/10.1002/pits.22114.

Goldberg, S. B., Wielgosz, J., Dahl, C., Schuyler, B., MacCoon, D. S., Rosenkranz, M., et al. (2015). Does the five facet mindfulness questionnaire measure what we think it does? Construct validity evidence from an active controlled randomized clinical trial. *Psychological Assessment.* https://doi.org/10.1037/pas0000233 No Pagination Specified.

Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources, 42*(4), 765–794. https://doi.org/10.1037/e722702011-001.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review Of Psychology, 60*, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530.

Graham, J. W., & Schafer, J. L. (1999). On the performance of multiple imputation for multivariate data with small sample size. *Statistical Strategies for Small Sample Research, 50*, 1–27.

Gregory, A., Hafen, C. A., Ruzek, E., Mikami, A. Y., Allen, J. P., & Pianta, R. C. (2016). Closing the racial discipline gap in classrooms by changing teacher practice. *School Psychology Review, 45*(2), 171–191. https://doi.org/10.17105/spr45-2.171-191.

Hare, T. A., Tottenham, N., Galvan, A., Voss, H. U., Glover, G. H., & Casey, B. J. (2008). Biological substrates of emotional reactivity and regulation in adolescence during an emotional go-nogo task. *Biological Psychiatry, 63*(10), 927–934. https://doi.org/10.1016/j.biopsych.2008.03.015.

Hargreaves, A. (1998). The emotional practice of teaching. *Teaching and Teacher Education, 14*(8), 835–854. https://doi.org/10.1016/s0742-051x(98)00025-0.

Hirshberg, M.J., Flook, L., Enright, R.D., & Davidson, R.J. Durable teacher reductions in implicit race bias through mindfulness-based skills training. Manuscript submitted for publication. 10.31231/osf.io/prmc9.

Hirshberg, M. J., Patsenko, E. G., Davidson, R. J., & Wilson-Mendenhall (2018). Behavioral measurement of mindfulness. In A. Bernstein (Ed.). *(Chair) advances in behavioral measurement of mindfulness. Symposia presented at the international symposium for contemplative research, phoenix, Arizona.*

Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology, 64*(2), 144–150. https://doi.org/10.1037/0021-9010.64.2.144.

Hue, M., & Lau, N. (2015). Promoting well-being and preventing burnout in teacher education: A pilot study of a mindfulness-based programme for pre-service teachers in Hong Kong. *Teacher Development, 19*(3), 381–401. https://doi.org/10.1080/13664530.2015.1049748.

Jamil, F. M., Downer, J. T., & Pianta, R. C. (2012). Association of pre-service teachers' performance, personality, and beliefs with teacher self-efficacy at program completion. *Teacher Education Quarterly, 39*(4), 119–138.

Jennings, P. A., Brown, J. L., Frank, J. L., Doyle, S., Oh, Y., Davis, R., et al. (2017). Impacts of the CARE for teachers program on teachers' social and emotional competence and classroom interactions. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000187.

Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research, 79*(1), 491–525. https://doi.org/10.3102/0034654308325693.

Jennings, P. A., Snowberg, K. E., Coccia, M. A., & Greenberg, M. T. (2011). Improving classroom learning environments by cultivating awareness and resilience in education (CARE): Results of two pilot studies. *Journal of Classroom Interaction,* 37–48.

Jerusalem, M., & Schwarzer, R. (1992). Self-efficacy as a resource factor in stress appraisal processes. In R. Schwarzer, & Schwarzer (Eds.). *Self-efficacy: Thought control of action* (pp. 195–213). Washington, DC, US: Hemisphere Publishing Corp (1992-97719-009).

Kabat-Zinn, J. (1982). An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: Theoretical considerations and preliminary results. *General Hospital Psychiatry, 4*(1), 33–47. https://doi.org/10.1016/0163-8343(82)90026-3.

Kabat-Zinn, J. (2013). *Full catastrophe living, revised edition: How to cope with stress, pain and illness using mindfulness meditation.* (Hachette UK).

Kang, Y., Gray, J. R., & Dovidio, J. F. (2014). The nondiscriminating heart: Lovingkindness meditation training decreases implicit intergroup bias. *Journal of Experimental Psychology: General, 143*(3), 1306. https://doi.org/10.1037/a0034150.

Kennedy, M. (1999). The role of preservice teacher education. *Teaching as the Learning Profession: Handbook of Policy and Practice,* 54–85.

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*. https://doi.org/10.3102/0034654315626800.

Kesebir, P., Gasiorowska, A., Goldman, R., Hirshberg, M. J., & Davidson, R. J. (2019). Emotional Style questionnaire: A multidimensional measure of healthy emotionality. *Psychological Assessment, 31*(10), 1234–1246. https://doi.org/10.1037/pas0000745.

Kim, L. E., Jörg, V., & Klassen, R. M. (2019). A meta-analysis of the effects of teacher personality on teacher effectiveness and burnout. *Educational Psychology Review*. https://doi.org/10.1007/s10648-018-9458-2.

Klingbeil, D. A., & Renshaw, T. L. (2018). Mindfulness-based interventions for teachers: A meta-analysis of the emerging evidence base. *School Psychology Quarterly, 33*(4), 501–511. https://doi.org/10.1037/spq0000291.

Klusmann, U., Kunter, M., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Teachers' occupational well-being and quality of instruction: The important role of self-regulatory patterns. *Journal of Educational Psychology, 100*(3), 702. https://doi.org/10.1037/0022-0663.100.3.702.

La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The classroom Assessment scoring system: Findings from the prekindergarten year. *The Elementary School Journal, 104*(5), 409–426. https://doi.org/10.1086/499760.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the implicit association test: IV. *Implicit Measures of Attitudes, 59*–102.

Lauermann, F., & König, J. (2016). Teachers' professional competence and wellbeing: Understanding the links between general pedagogical knowledge, self-efficacy and burnout. *Learning and Instruction, 45*, 9–19. https://doi.org/10.1016/j.learninstruc.2016.06.006.

Levinson, D. B., Stoll, E. L., Kindy, S. D., Merry, H. L., & Davidson, R. J. (2014). A mind you can count on: Validating breath counting as a behavioral measure of mindfulness. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.01202.

Lutz, A., Jha, A. P., Dunne, J. D., & Saron, C. D. (2015). Investigating the phenomenological matrix of mindfulness-related practices from a neurocognitive perspective. *American Psychologist, 70*(7), 632. https://doi.org/10.1037/a0039585.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *Maslach burnout inventory manual.* Mountain View, CA: CPP. Inc., and Davies-Black.

McCann, T. M., & Johannessen, L. R. (2004). Why do new teachers cry? *The clearing house. A Journal of Educational Strategies, Issues and Ideas, 77*(4), 138–145. https://doi.org/10.3200/TCHS.77.4.138-145.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119. https://doi.org/10.3102/0013189x09332374.

Quirin, M., Kazén, M., & Kuhl, J. (2009). When nonsense sounds happy or helpless: The implicit positive and negative affect test (IPANAT). *Journal of Personality and Social Psychology, 97*(3), 500. https://doi.org/10.1037/a0016063.

Rinpoche, P. (1993). *The heart treasure of the enlightened ones: The practice of view, meditation, and action.* Shambhala Publications.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica,* 417–458. https://doi.org/10.1111/j.1468-0262.2005.00584.x.

Roeser, R. W., Schonert-Reichl, K. A., Jha, A., Cullen, M., Wallace, L., Wilensky, R., et al. (2013). Mindfulness training and reductions in teacher stress and burnout: Results from two randomized, waitlist-control field trials. *Journal of Educational Psychology, 105*(3), 787–804. https://doi.org/10.1037/a0032093.

Roeser, R. W., Skinner, E., Beers, J., & Jennings, P. A. (2012). Mindfulness training and teachers' professional development: An emerging area of research and practice. *Child Development Perspectives, 6*(2), 167–173. https://doi.org/10.1111/j.1750-8606.2012.00238.x.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys, Vol.81.* John Wiley & Sons.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57*(6), 1069.

Ryff, C. D., & Keyes, C. L. (1995). The structure of psychological well-being revisited. *Journal Of Personality And Social Psychology, 69*(4), 719–727. https://doi.org/10.1037/0022-3514.69.4.719.

Sandilos, L. E., DiPerna, J. C., & Investigators, T. F. L. P. K. (2014). Measuring quality in kindergarten classrooms: Structural analysis of the classroom Assessment scoring system (CLASS K–3). *Early Education & Development, 25*(6), 894–914. https://doi.org/10.1080/10409289.2014.883588.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.2307/1175860.

Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods, 6*(4), 317. https://doi.org/10.1037/1082-989x.6.4.317-329.

Smith, M. W., Brady, J. P., & Anastasopoulos, L. (2008). *Early language and literacy classroom observation pre-K (ellco pre-K) user's guide.* Paul H. Brookes.

Team, R. C. (2014). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing 2012. ISBN 3-900051-07-0.

Thrangu, K., & Thrangu, R. (2004). *An ocean of the ultimate meaning: Teachings on mahamudra: A commentary on wangchuk dorje's ngedön gyamtso.* Shambhala Publications.

Vago, D. R., & Silbersweig, D. A. (2012). Self-awareness, self-regulation, and self- transcendence (S-ART): A framework for understanding the neurobiological mechanisms of mindfulness. *Frontiers in Human Neuroscience, 6.* https://doi.org/10.3389/fnhum.2012.00296.

Van Dam, N. T., Earleywine, M., & Danoff-Burg, S. (2009). Differential item function across meditators and non-meditators on the five facet mindfulness questionnaire. *Personality and Individual Differences, 47*(5), 516–521. https://doi.org/10.1016/j.paid.2009.05.005.

Van Dam, N. T., van Vugt, M. K., Vago, D. R., Schmalzl, L., Saron, C. D., Olendzki, A., et al. (2017). Mind the hype: A critical evaluation and prescriptive agenda for research on mindfulness and meditation. *Perspectives on Psychological Science.* https://doi.org/10.1177/1745691617709589.

Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry, 14*(3), 270–277. https://doi.org/10.1002/wps.20238.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063. https://doi.org/10.1037/0022-3514.54.6.1063.

Weng, H. Y., Fox, A. S., Shackman, A. J., Stodola, D. E., Caldwell, J. Z., Olson, M. C., et al. (2013). Compassion training alters altruism and neural responses to suffering. *Psychological Science, 24*(7), 1171–1180. https://doi.org/10.1177/0956797612469537.

White, I. R., Horton, N. J., Carpenter, J., Pocock, S. J., & others (2011). Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ, 342,* d40. https://doi.org/10.1136/bmj.d40.