THE EFFICACY OF INTERVENTIONS FOR REDUCING TEACHER BURNOUT: A META-ANALYSIS

A Dissertation

by

JADE GAIL KESTIAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Cynthia Riccio
Committee Members,	Constance Fournier
	Jean Madsen
	Christopher Thompson
Head of Department,	Shanna Hagan-Burke

May 2020

Major Subject: School Psychology

ABSTRACT

The present study is a systematic meta-analytic study of 30 primary sources investigating the effectiveness of interventions to treat teacher burnout. We systematically searched and screened over 6,000 records for studies to include. We computed summary effect sizes for each facet of burnout and total burnout symptoms at post-intervention and follow-up. We conducted 23 separate meta-regressions for potential moderators found to influence burnout our related constructs in prior literature. Results indicate that interventions to prevent and reduce teacher burnout do appear to have some evidence of effectiveness. Summary effect sizes at post-intervention were small for Emotional Exhaustion *(k* =25, *g* = 0.16, 95% CI, [0.09, 0.23], *p* <.001), Personal Accomplishment (*k* =19, *g* = 0.21, 95% CI, [0.06; 0.35], *p* < .01), and Total Burnout (*k* = 6 *g* = 0.35, 95% CI, -0.02, 0.72, *p* = .06). The summary effect size for Cynicism/Depersonalization was trivial at post-intervention (*k* =17, *g* = 0.14, 95% CI, [0.01, 0.27], *p* = .037). Summary effect sizes at follow-up were moderate for Emotional Exhaustion (*g* = 0.52, 95% CI [-0.03, 1.07], *p* = .056), and small for Cynicism/ Depersonalization, and Personal Accomplishment. Sample sizes were small for follow-up data; however, results can be cautiously interpreted as indicating that intervention effects are maintained over time. Mindfulness interventions appear to have more, and stronger evidence compared to behavioral interventions for the purposes of reducing burnout symptoms. Minutes of direct contact moderated effect sizes for Personal Accomplishment, suggesting potentially important implications for structuring interventions. We discuss interpretations and implications of these results, as well as limitations and future directions for teacher burnout intervention research.

DEDICATION

This work is dedicated to all the teachers who empower youth to become the best versions of themselves. This dissertation is for you, in the hopes that each school system can use the most up-to date high-quality evidence to make small and large changes to support you and the incredibly important work you do. You make a difference in so many lives. I hope you can obtain the resources you need to care for yourselves and one another, so that your lives can be full and peaceful, and so that your compassion can shine into the lives of the children in your schools, their families, and your co-workers. I hope that you find the mentors, the advice, the support, and the calm moments that you need to thrive in your careers and your lives.

ACKNOWLEDGEMENTS

This dissertation has been a labor of love, sweat, some small tears, and perseverance that was loaned to me from many incredibly talented, hardworking, and compassionate individuals. I am exceptionally grateful to every one of you.

I am eternally indebted to Cynthia Riccio, Ph.D. for her guidance and mentorship, for her help in refining the coding sheet, for coding 30% of the included studies in record time, and for gently dragging me across the finish line. Dr. Riccio, you are a true superhero of the School Psychology program at Texas A&M and the field in general; your legacy will continue long after you retire, not only in your work, but also in the careers and work of your countless students and mentees.

Thank you, Constance Fournier, Ph.D., for your mentorship, encouragement, and words of wisdom. My house, like yours, has never been cleaner than during this past year of dissertation work. This work, while not world-changing, has forever made an imprint on my psyche such that I know I can accomplish the goals I set out to accomplish—even if it takes 6+ years. Thank you for reminding me to let go and get it done.

Christopher Thompson, Ph.D., I am immensely grateful for the large swath of time you took to discuss our work during the final stages. Thank you for all your assistance walking me through using R via Skype, for answering my myriad questions, and for your honest feedback. You have helped me grow as a researcher in ways I previously thought I could not grow.

To Jean Madsen, Ph.D., thank you for providing your time, your perspective, and your feedback from proposal through defense. Your diverse perspective as an outside department committee member is very much appreciated and needed.

I also want to thank Dr. Bowman-Perrott for your contributions to this work from its early stages. Your advice and recommendations were incredibly meaningful and helpful in re-working the initial proposal into the final product.

Finally, I would like to thank my family, including the people who raised me, and the newer, chosen family in my life. Thank you for your love, your generosity, your words of encouragement, and for limiting the number of times you all asked, "when will you be done, again?" To my husband C.J., I am incredibly fortunate and grateful to have you as a life partner. Thank you for your unconditional love, for making me lentil soup, for reminding me to eat, for taking on more than your fair share of the household responsibilities during the final stretch, and for continuing to encourage and cheer me on--even when I rant about the various methods for estimating tau squared (τ^2) for far longer than anyone would care to listen. You are my favorite human and the first person on my gratitude list every day.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Cynthia Riccio, Ph.D., Constance Fournier Ph.D., Christopher Thompson, Ph.D., and Shanna Hagan-Burke of the Department of Educational Psychology, and Jean Madsen, Ph.D. of the Department of Educational Administration and Human Resource Development.

The data analyzed as a part of this meta-analysis was coded by the student and 30% of the studies were also coded by Cynthia Riccio, Ph.D. The analyses depicted in Chapter IV were guided in part by Christopher Thompson, Ph.D. of the Department of Educational Psychology. All other work conducted for this dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by several scholarships and fellowships from Texas A&M University, including the Graduate Merit Fellowship and the 2018-2019 Internship Scholarship Award. Additional support was provided by teaching assistantships, a year-long, full-time, paid internship at Travis County Juvenile Probation Department, and family funds.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Significance of Problem

Traditional conceptualizations of burnout define it as a syndrome where professionals feel exhausted and cynical about their work (Hakanen et al., 2006; Maslach et al., 2001). Some researchers believe that helping professionals or individuals who provide human services are at an increased risk of burnout compared to other professions (Maslach & Jackson, 1981); some suggest teachers appear to be at especially high risk (De Heus & Diekstra, 1999).

Burnout is associated with health complaints such as musculoskeletal pain (Armon et al., 2010), depression (Bianchi et al., 2013; Melamed et al., 2006), and coronary heart disease (Hallman et al., 2003), as well as cardiovascular disease among men and musculoskeletal disorders among women (Honkonen et al., 2006). A review of the literature suggests that burnout and vital exhaustion increase risk of myocardial infarction, stroke, sudden cardiac death, Type 2 diabetes, and impairment of reproductive function (Melamed et al., 2006).

Interest in teacher wellbeing and how teacher wellbeing is related to student outcomes and school ecology is increasing (Jennings & Greenberg, 2009; Roffey, 2012). Among teachers, burnout appears to be related to student behavior, academic functioning, and school performance (Briner & Dewberry, 2007; Hastings & Bham, 2003; Kokkinos, 2007; Ruble & McGrew, 2013). Burnout symptoms have been found to negatively relate to Individualized Education Plan (IEP) goal attainment for children with autism (Ruble & McGrew, 2013). That is, the students whose teachers exhibit higher levels of burnout symptoms are less likely to attain their IEP goals.

Burnout symptoms also relate to teacher's perceptions of student emotional and behavioral problems such that teachers with higher burnout symptoms are more likely to rate their students as having higher levels of emotional and behavioral problems (Hastings & Bham, 2003; McLean et al., 2019). The cross-sectional and self-report nature of most research on student outcomes and burnout precludes conclusions about the direction of these relationships, whether relationships are bidirectional, or whether the correlations are due to confounding variables. These associations may result from an introduction of bias into social-emotional and behavioral screening tools by the burned-out teachers, or could represent actual differences in behavior and emotional problems that are precipitated by teachers' low self-efficacy and burnout symptoms, or both (McLean et al., 2019).

Broader measures of teacher wellbeing, including job enjoyment, also have been found to relate to measures of school performance, such as standardized test scores; however, these results only suggest associations at the school level (Briner & Dewberry, 2007). While perceptions of misbehavior appear to be positively related to burnout symptoms, other longitudinal research that utilized student discipline records uncovered unexpected relationships between burnout symptoms and out of school suspensions, in that teachers who had higher levels of burnout were 32% less likely to have students who received a suspension by the spring (Pas et al., 2010). It is possible that burned-out teachers are less likely to engage in discipline referrals despite increased perceptions of behavior problems due to their pessimism and lack of classroom management efficacy (Pas et al., 2010).

Teacher burnout is especially concerning to school systems due to its link to teacher organizational commitment, job satisfaction, and turnover intentions (Alarcon, 2011). While teacher production has declined in recent years, it has increased overall since the 1980's (Cowan

et al., 2016; Sutcher et al., 2016). Despite the overall increase in production, teacher shortages in many subjects and settings, such as in science, technology, engineering, and mathematics (STEM), in special education (SPED), and in disadvantaged schools, are still a very real issue (Cowan et al., 2016; Goldhaber et al., 2015). The single largest factor affecting teacher shortages is not production, rather, it is teacher turnover or attrition (Sutcher et al., 2016). The United States has a high rate of teacher attrition (8%) compared to high achieving areas including Finland, Singapore, and Ontario, Canada (Darling-Hammond, & Rothman, 2011). Most teachers who choose to leave their schools or profession cite dissatisfaction in their current placement as a major reason for leaving (Sutcher et al., 2016). This dissatisfaction is conceptually related to the construct of burnout.

Statement of Problem

Teacher burnout is a problem that impacts teachers, school systems, and students. Much research on burnout has been cross-sectional, though approaches for preventing teacher burnout and increasing teacher resilience are receiving increased attention in recent decades (Beltman et al., 2011; Chang, 2009). The experimental research that does exist is beginning to be synthesized in the literature. Iancu et al. (2018) conducted a meta-analysis to determine the effectiveness of interventions aimed at reducing teacher burnout and found that, overall interventions had a small, statistically significant mean effect size. Iancu et al. (2018), however, did not examine some relevant potential moderators, such as teacher race and/or ethnicity, years' experience teaching, or teacher age. Further, the search procedure used by Iancu et al. (2018) has not been replicated with additional keywords to ensure their review captured all possible records. Zarate et al. (2019) also conducted a meta-analysis on teacher mental health outcomes, including burnout, for mindfulness training specifically; however, they did not include other types of interventions.

Purpose of Study

The purpose of this study is to perform a systematic meta-analysis to expand on what is currently known about interventions to prevent and treat teacher burnout. The present study has expanded search criteria compared to Iancu et al. (2018) and will measure study quality in a different way. This study also will examine additional potential moderators of intervention effectiveness. These nine potential moderators include factors associated with burnout or associated with related constructs such as turnover or teacher shortages.

Like Iancu et al. (2018), this study uses a random effects model, presents confidence intervals for each study effect size, and examines potential moderators using meta-regression. Unlike Iancu et al. (2018), the present study included additional key words within the search, uses adjusted standardized mean difference effect sizes to summarize intervention effects, examined study quality in a different way, and examined some different potential moderators. While Iancu et al. (2018), did not specify their method for estimating tau squared (τ^2) the present study uses both the Sidik-Jonkman (SJ) method and Restricted Maximum Likelihood (REML) as described in the results.

It is beyond the scope of the present study to examine retention outcomes for experimental mentoring programs; however, a meta-analytic research examining mentoring's effect on retention outcomes is needed (Ingersoll & Kralik, 2004).

Research Questions

1. What are the average weighted effect sizes for interventions to prevent teacher burnout, Emotional Exhaustion, Cynicism/Depersonalization, and Personal Accomplishment? We hypothesized that overall, interventions will produce a small effect.

2. For those studies that included follow-up data, were effects maintained over time? We hypothesized that only trivial differences will remain at follow-up.

3. Do different types of interventions have stronger effects? We hypothesized that mindfulness interventions targeting teachers will have larger effects than classroom management or other behavioral interventions.

4. What teacher and school characteristics moderate intervention effectiveness? We hypothesized that many factors, including length of intervention, use of homework, work experience, age, sex, and race may moderate effectiveness. The following hypotheses were considered:

    a) We hypothesized that studies with interventions implemented over longer periods of time would produce larger effect sizes for each facet of burnout and for total burnout scores.

    b) We hypothesized that interventions that incorporated homework would produce larger effects for each facet of burnout and total burnout scores.

    c) We hypothesized that samples with a lower mean for years of work experience would see a greater reduction in burnout symptoms for all three facets of burnout and total burnout scores.

    d) We hypothesized that samples with a lower mean age would see greater reductions in Emotional Exhaustion and Cynicism/Depersonalization compared to samples with lower proportions of younger teachers.

e) We hypothesized that samples with relatively higher proportions of female teachers would see a greater reduction in Emotional Exhaustion compared to samples with lower proportions of female teachers.

f) We hypothesized that samples with higher proportions of male teachers would see a greater reduction in Cynicism/Depersonalization compared to samples with lower proportions of male teachers.

g) We hypothesized that samples with higher proportions of White teachers would show greater reduction in burnout symptoms across domains compared to samples with lower proportions of White teachers.

h) We hypothesized that samples conducted in schools with lower proportions of students with free and reduced lunch will show greater reductions in burnout symptoms across domains.

Definitions of Terms

Burnout

Burnout is a mental health construct traditionally conceptualized as having three components: Emotional Exhaustion, low Personal Accomplishment, and Cynicism/ Depersonalization (Maslach & Jackson, 1981; Maslach et al., 2001). Emotional exhaustion often considered to be the core component of burnout (Brouwers & Tomic, 2000). Burnout is not the same as attrition, turnover, or retention; however, burnout is conceptually related to job dissatisfaction, and thus may be a cause of attrition/turnover (Sutcher et al., 2016). Individuals who are "burned-out" have not necessarily left the profession; rather, they feel exhausted, cynical, and unaccomplished in their profession. Some recent research found that both Emotional

Exhaustion and Cynicism/Depersonalization load on the same general factor as depression and anxiety symptoms (Schonfeld et al., 2019a).

Emotional Exhaustion

Emotional exhaustion refers to feelings of aversive chronic stress, typically due to demands of the classroom as well as a lack of support and coping mechanisms (Maslach & Jackson, 1981; Maslach et al., 2001). Schonfeld et al., (2019a) suggest that Emotional Exhaustion scale of the MBI is more highly related to depression and anxiety symptoms than it is to Personal Accomplishment and Cynicism/Depersonalization.

Low Personal Accomplishment

Low Personal Accomplishment involves evaluating yourself negatively and feeling dissatisfied with work-related accomplishments or feeling as though you have not accomplished anything (Maslach & Jackson, 1981; Maslach et al., 2001). Some argue that Personal Accomplishment is distinct from self-efficacy (Bandura, 1977; Brouwers & Tomic, 2000), while others use the terms interchangeably (Boersma & Linblom, 2009). For the purposes of this paper, Personal Accomplishment was viewed as distinct from self-efficacy.

Depersonalization

Depersonalization refers to having a cynical attitude or feeling cold or distant towards students (Maslach & Jackson, 1981; Maslach et al., 2001). Some research refers to Cynicism/Depersonalization as cynicism, and thus, the present study will refer to this construct as Cynicism/Depersonalization. Schonfeld et al. (2019a) found that approximately half of the variance of the Depersonalization scale of the MBI loads onto a general non-specific psychological distress (NPD) factor, and the other half loads onto a separate Depersonalization factor.

Self-efficacy

Self-efficacy describes individuals' beliefs about their ability to use cognitive resources, mobilize motivation, and create plans of action to accomplish a given task within a situation (Bandura, 1977). Different types of self-efficacy are defined based on the context in which they are relevant. Teaching self-efficacy is thus relevant to teachers and teaching and can be broken down further into more specific types of self-efficacy (e.g. classroom management self-efficacy: one's belief about their ability to manage behavior appropriately in the classroom). Self-efficacy is strongly related to Personal Accomplishment; however, it is typically considered a separate construct. (Shoji et al., 2016).

Mindfulness

Mindfulness is considered both an intentional practice as well as an outcome within the literature. One prominent definition defines mindfulness practice as paying attention, in a particular way, on purpose, in the present moment, non-judgmentally (Kabat-Zinn, 1990). Operational definitions of the quality of mindfulness include emotional and cognitive components, such as awareness of sensations, awareness of thoughts, awareness of emotions, self-regulation of attention, as well as openness and acceptance of these thoughts and feelings with calmness, non-reactivity, and non-judgement (Khoury et al., 2013).

Mindfulness relates to the quality of one's intimate relationships, perhaps by improving the quality of emotional interactions, allowing partners to regard each other more nonjudgmentally (Wachs & Cordova, 2007). Taken together, these definitions suggest the quality of being "mindful" overlaps with aspects of the core social-emotional competencies defined by the Collaborative for Academic and Social Emotional Learning (CASEL), including self-

awareness, self-management, and social-awareness; further, mindfulness may also be related to relationship skills (CASEL, 2013; CASEL, 2015).

Mindfulness interventions can take many forms; common approaches to teaching mindfulness practices include teaching deep breathing, guided or silent mindfulness meditation, yoga, or some combination of these activities. While approaches to teaching or improving mindfulness vary, practicing mindfulness generally means developing self-awareness and non-judgment of inner experiences.

Implications for Practice

The results of this study will inform recommendations for interventions to prevent and treat teacher burnout. The study will describe the types of interventions that have evidence of efficacy, the summary standardized mean difference effect sizes for reductions in burnout subscales and total scores, and which study characteristics appear to moderate intervention effectiveness for subscales and total scores. Further, this study will compare the present methods and results to those obtained by Iancu et al. (2018) in their similar meta-analysis of teacher burnout interventions. Finally, this study will yield valuable information regarding what additional research needs to be conducted on interventions for teacher burnout.

CHAPTER II

LITERATURE REVIEW

History of Studying Teacher Burnout

Burnout has been a topic of research since the 1970's (Freudenberger, 1974; Freudenberger 1975; Maslach, 1978), and the construct of burnout had been a topic of literature and social commentary long before it was studied systematically (Maslach et al., 2001). Maslach et al. (2001) break down the history of research on burnout into two distinct phases: the pioneering phase and the empirical phase.

Pioneering Phase

The pioneering phase helped to define burnout as a construct and demonstrate burnout as a relatively common workplace response. During this phase, Maslach and others used interviews, case-studies, and observations within social service professions to study and define the construct of burnout (Maslach et al., 2001). From this research, Maslach et al. (2001) developed a tripartite model of burnout, which defines burnout as including Emotional Exhaustion, Cynicism/Depersonalization, and Personal Accomplishment. This definition of burnout differs from colloquial uses of the term "burnout" in that it is distinct from constructs such as retention, attrition, and turnover. The terms attrition and turnover refer to the rate at which professionals leave a given profession. Retention refers to the rate at which professionals remain in a profession. Burnout symptoms can lead to attrition or turnover; however, burnout is not synonymous with attrition nor turnover. For the purposes of this study, burnout was defined as a mental health construct comprised of Emotional Exhaustion, Cynicism/Depersonalization, and lack of Personal Accomplishment.

Empirical Phase

The empirical phase began in the 1980's, when research on burnout shifted to become more quantitative. Researchers began developing questionnaires and surveys to use with larger samples. Burnout measures, such as the Maslach Burnout Inventory (MBI; Maslach & Jackson, 1981) began to emerge and gather evidence of yielding reliable and valid data. Research on burnout also began to expand to populations beyond social-service workers and began examining how facets of burnout related to other factors, such as personality and workplace factors.

By the 1990's, burnout was a topic of international research, and the MBI (Maslach & Jackson, 1981) had been translated into several other languages, including German, Dutch, and French, and translations had been studied psychometrically (Maslach et al., 2001). Currently, intervention research (Iancu et al., 2018; West et al., 2016; Zarate et al., 2019) and meta-analyses related to burnout are increasingly being published to fill gaps in the literature (Alarcon, 2011; Aloe et al., 2014; Brewer & Shapard, 2004; Edmonson & Thompson, 2000; Halbesleben, 2006; Iancu et al., 2018; Ng & Feldman, 2012; Purvanova & Muros, 2010; Reichl et al., 2014; Shin et al., 2014; Shoji et al., 2016; Swider & Zimmerman, 2010; West et al., 2016; Zarate et al., 2019). Further, some recent research critiques the construct of burnout, particularly its discriminant validity and high correlations with neuroticism and depression measures, and these researchers suggest that attention should be shifted away from studying burnout and toward studying depression (Schonfeld et al., 2019b). Schonfeld et al. (2019b) suggest that the label "burnout" may minimize the severity of the condition and that labeling "burnout" as "depression," along with studying depression as an outcome in occupational health research, may increase help-seeking behavior.

Burnout in Educational Settings

The term "teacher burnout" appears in the literature in the 1980's. Seidman and Zager (1986) developed The Teacher Burnout Scale, to assess burnout in teachers specifically; this scale is now used infrequently. Later, Maslach, Jackson, and Schwab (1996) modified the original MBI (Maslach & Jackson, 1981) to create the Maslach Burnout Inventory Educator Survey (MBI-ES). The first obvious examples of intervention research on preventing and treating teacher burnout began in the late 1990's and early 2000's (Anderson et al., 1999; Neves de Jesus & Conboy, 2001). Burnout intervention research has increased substantially since 2010. Within the teacher burnout literature, interventions are almost exclusively aimed at individual teachers as opposed to school leadership or the organizational culture of schools. Recent research on teacher burnout includes several meta-analyses examining relationships with other factors and supports several of the more prominent theoretical frameworks used to conceptualize burnout (Aloe et al., 2014; Edmonson & Thompson, 2000).

Theoretical Considerations

Several theoretical frameworks have been used to study burnout, including Job Demands--Resources Model (Hakanen et al., 2006; Fernet et al., 2013), Self-Determination Theory (Deci & Ryan, 2008; Fernet et al., 2012), Social Cognitive theory (Aloe et al., 2014; Bandura, 1977; Brown, 2012), and Conservation of Resource Theory (Alarcon, 2011). These theoretical frameworks work in concert to explain the process of burnout (Fernet et al., 2013).

Job Demands-Resources Model

The Job Demands-Resources Model of burnout (Demerouti et al., 2001) posits that job demands and job resources are differentially related to aspects of burnout and other outcomes, such that job demands are more closely related to exhaustion and the lack of job resources is

more closely related to disengagement from work tasks, and Cynicism/Depersonalization in occupations where job tasks include working with people. Job demands are the physical, social, and organizational aspects of an occupation that require sustained effort (Demerouti et al., 2001). This sustained effort is theorized to link more directly with the exhaustion dimension of burnout and the process of job demands affecting burnout is considered an energetical process (Demerouti et al., 2001; Hakanen et al., 2006). Job resources are the physical, psychological, social, or organizational aspects of the job that may help one achieve work goals, reduce job demands or the physical or psychological costs of those demands, or stimulate personal development (Demerouti et al., 2001). Job resources are thought to relate to a motivational process, such that lack of resources hampers motivation, leading to disengagement or Cynicism/Depersonalization (Bakker & Demerouti, 2007; Demerouti et al., 2001; Hakanen et al., 2006). The literature focuses primarily on the external resources, such as job control, administrative leadership, deficient equipment, school policies, school climate, potential for qualification, participation in decision-making, and task variety (Demerouti et al., 2001; Fernet et al., 2013). In addition to being theorized as more closely related to engagement or lack of Cynicism/Depersonalization, job resources are also thought to at least partially buffer the negative effects of job demands (Bakker & Demerouti, 2007).

A review of the literature on the Job Demands-Resources Model and burnout suggests that job demands are primarily related to exhaustion and job resources are primarily related to Personal Accomplishment and Cynicism/Depersonalization, though some overlap exists (Fernet et al., 2013). With respect to teacher burnout specifically, Hakanen et al. (2006) found evidence of both the energetical and motivational processes at work. They also found that burnout symptoms appeared to mediate the effect of job demands (e.g. overload, student behavior

problems, and physical environment) on poor health. Hakanen et al. (2006) further found that work engagement mediated the effects of job resources (e.g. job control, information, supervisor support, social climate) on organizational commitment, and that that the effect of lack of resources on low engagement was mediated by burnout. Evidence of the energetical process was more prominent (Hakanen et al. (2006). Further, the perception of a supportive organizational climate appears to be a protective factor for teacher burnout (Lavian, 2012). The cross-sectional and self-report nature of these data preclude any firm conclusions about the true relationships between each construct.

Self-Determination Theory

Some research on burnout has focused on incorporating Self-Determination Theory within the Job Demands-Resource Model (Fernet et al., 2013; Fernet et al., 2012). Self-Determination Theory (Deci & Ryan, 2008) suggests autonomous motivation, that is, the experience of choice in initiating behavior, is essential for ideal functioning. Autonomous motivation, or intrinsic motivation, has a strong theoretical link with self-efficacy (Deci & Ryan, 2008), an important variable in predicting burnout (Aloe et al., 2014; Brouwers & Tomic, 2000; Shoji et al., 2016). Researchers who use Self-Determination Theory to study burnout within the Job Demands-Resources Model suggest it is difficult to completely differentiate the energetic and motivational processes proposed by Job Demands-Resources Model because both processes may be related to autonomous motivation and self-efficacy (Fernet et al., 2013; Fernet et al., 2012). It may be that a lack of resources or high job demands play a role in burnout because they erode teacher's self-efficacy.

Social Cognitive Theory

The concept of self-efficacy is derived from Social Cognitive Theory (Bandura, 1977). Self-efficacy, broadly defined, is an individual's belief about their ability to use motivation, cognitive resources, and plans of action to accomplish a specific task within a specified context (Bandura, 1977). In this way, self-efficacy is a major determinant of motivational processes (Bandura, 1977; Schwarzer & Hallum, 2008). Research supports the notion that self-efficacy is important for teacher resilience and burnout (Aloe et al., 2014; Beltman et al., 2011; Brown, 2012). Self-efficacy has a negative relationship with burnout (Aloe et al., 2014; Brown, 2012) and is considered a mediator between work stress and burnout (Yu et al., 2015).

Teacher self-efficacy is currently considered a multi-dimensional construct describing the extent to which a teacher believes they can teach even difficult/unmotivated students (Tschannen-Moran & Woolfolk-Hoy, 2001; Skaalvik & Skaalvik, 2007). Types of teacher self-efficacy within the literature vary and may include instructional efficacy, engagement efficacy, and classroom management efficacy (Tschannen-Moran & Woolfolk-Hoy, 2001), or efficacies related to instruction, adapting education to fit students, motivating students, keeping order and discipline, cooperating with colleagues and parents, and coping with changes and challenges (Skaalvik & Skaalvik, 2007). General self-efficacy and classroom management self-efficacy have both been studied and found to relate to teacher burnout (Aloe et al., 2014; Brown, 2012).

Conservation of Resource Theory

Conservation of Resource Theory is closely related to Job Demands-Resources Theory, though this theory defines resources differently, and conceptualizes job demands as a potential threat to resources (Hobfoll, 1989). Conservation of Resource Theory suggests people aim to acquire and maintain resources (Hobfoll, 1989). Within this framework, stress is defined as a

reaction to the environment due to one of three things: a threat of a net loss of resources, an actual net loss of resources, or a lack of resource gain following an investment of resources (Hobfoll, 1989). Resources are defined comprehensively as "objects, personal characteristics, conditions or energies that are valued by the individual or that serve as a means for attainment of these objects, personal characteristics, or energies" (Hobfoll, 1989, p. 516). Further, the model predicts individuals will attempt to minimize their net loss of resources when confronted with stress, potentially by limiting their engagement (Hobfoll, 1989; Ng & Feldman, 2012). Teacher burnout is easily conceptualized using this model, with the demands of teaching taxing resources such as time and emotional energy, often without a substantial return on this investment (Alarcon, 2011). The diminished resources lead to stress, which triggers disengagement, and can lead to a downward spiral of maladaptive coping, culminating in burnout symptoms (Alarcon, 2011).

In addition to support for Job Demands-Resources Model, which can be interpreted as support of Conservation of Resources Theory (Fernet et al., 2013; Fernet et al., 2012; Hakanen et al., 2006), other empirical investigation of burnout, and other constructs, has yielded some support for Conservation of Resources theory and supports its utility for studying burnout (Alarcon, 2011; Ng & Feldman, 2012). Meta-analytic findings suggest that all three aspects of burnout are correlated with role ambiguity, role conflict, and workload, supporting the notion that job demands tax resources and lead to stress and burnout (Alarcon, 2011). Similarly, all three dimensions of burnout were related to control and autonomy, suggesting these resources may have a protective function (Alarcon, 2011). In a meta-analysis examining employee voice behavior, researchers found that voice behavior, that is, employee's constructive, change-oriented communication, was negatively related to employee stress, supporting the notion of

resource conservation when under stress (Ng & Feldman, 2012). The limitations of this cross-sectional research preclude any firm conclusions about causation or whether other variables play a role in these relationships (Alarcon, 2011; Ng & Feldman, 2012).

Review of Existing Literature

One purpose of the present meta-analysis is to determine whether certain factors moderate intervention effectiveness, including the type of intervention. In determining what potential moderators to examine, it is important to consider all the variables known to impact burnout symptoms, as well as the types of interventions used to treat burnout and how they are theoretically related to burnout. Here, we review previous meta-analyses on burnout (Alarcon, 2011; Aloe et al., 2014; Brewer & Shapard, 2004; Edmonson & Thompson, 2000; Halbesleben, 2006; Ng & Feldman, 2012; Purvanova & Muros, 2010; Reichl et al., 2014; Shin et al., 2014; Shoji et al., 2016; Swider & Zimmerman, 2010), and other literature to support the inclusion of potential moderators in the present study.

Individual Factors

Studies using hierarchical modeling to parse out the variance in burnout symptoms have shown that most of the variability in scores is accounted for at the individual teacher level as opposed to the school level, supporting the theory that burnout is closely related to individual differences (McCarthy et al., 2009; Ullrich, Lambert, & McCarthy, 2012). Individual differences such as big five personality factors and perceptions of one's job demands and resources, appear related to all three facets of burnout (McCarthy et al., 2016; Swider & Zimmerman, 2010).

Personality

A meta-analysis examining personality and burnout found that all big five personality factors were related to each dimension of burnout; the factors within the five factor model

explained 33% of the variance in Emotional Exhaustion, 20% of the variance in Cynicism/Depersonalization, and 27% of the variance in Personal Accomplishment (Swider & Zimmerman, 2010). Neuroticism was most strongly related to Emotional Exhaustion (r = .52) and Cynicism/Depersonalization (r = .42) and had the second strongest relationship with Personal Accomplishment (r = -.38), after extraversion. Extraversion had the strongest relationship with Personal Accomplishment (r = .41); individuals low on extraversion also had lower Personal Accomplishment (Swider & Zimmerman, 2010).

Coping

Emotional responses and coping also appear related to burnout. Again, meta-analytic findings suggest emotional responses to stressful events have a strong influence on burnout, and that these emotional responses have moderate correlations with personality mediator variables, suggesting teachers' emotional responses are related to their personality traits (Montgomery & Rupp, 2005). Further, problem focused and emotion focused coping relate to burnout; Personal Accomplishment had a stronger relationship with problem focused coping compared to the other dimensions of burnout, and Emotional Exhaustion and Cynicism/Depersonalization had a stronger relationship with emotion-focused coping compared to Personal Accomplishment (Shin et al., 2014).

Sex

Meta-analyses also have found sex differences in burnout symptoms (Purvanova & Muros, 2010; Reichl et al., 2014). Women, overall, are more likely to experience Emotional Exhaustion compared to men, whereas men are more likely to exhibit Cynicism/ Depersonalization compared to women (Purvanova & Muros, 2010). Interestingly, differences did not appear to differ significantly in male typed vs. female-typed occupations; this suggests

that sex or gender, and not the occupation itself, was related to these differences (Purvanova & Muros, 2010). It is unclear how other factors related to gender and gender identity may impact burnout; there is a dearth of research for transgender and non-binary teacher populations. For example, none of the studies included in the present meta-analysis provided information on gender, only binary sex categorization. Another meta-analysis examining the relationship between work-nonwork conflict and burnout found that studies with higher proportions of women tended to have stronger correlations between work-non-work conflict and Emotional Exhaustion, whereas studies with higher proportions of men had stronger correlations between work-nonwork conflict and Cynicism/Depersonalization (Reichl et al., 2014).

Age and Teaching Experience

Age and experience may also be important variables in predicting burnout, or response to intervention. Teachers tend to be at a higher risk of burnout early in their careers (Gavish & Friedman, 2010). Meta-analytic findings suggest that age and years of experience have a negative relationship to symptoms of Emotional Exhaustion for teachers overall (Brewer & Shapard, 2004). A meta-analysis by Edmonson and Thompson (2000) found that less experienced special education teachers tended to have slightly higher levels of burnout, for all three facets of burnout, and that young age was associated with somewhat higher levels of Emotional Exhaustion and Cynicism/Depersonalization; however, age and experience explained a relatively small proportion of the variance in burnout scores (Edmonson & Thompson, 2000).

Race and Ethnicity

Little is known about how race and ethnicity affect burnout symptoms specifically; however, some research has examined how race and/or ethnicity is associated with constructs related to burnout, such as teacher turnover or attrition (Achinstein et al., 2010; Ingersol & May,

2016). Teachers of color are at an increased risk of turnover compared to White teachers (Achinstein et al., 2010; Ingersol & May, 2016). This difference may be due in part to the fact that low-income schools with high youth of color populations are more likely to have teachers of color and these schools tend to have higher rates of turnover (Achinstein et al., 2010; Ingersol & May, 2016). There is also some evidence that non-White populations have poorer health outcomes and marginally higher levels of depressive symptoms compared to White individuals (Menselson et al., 2008; Pinquart & Sörensen 2005; Tarricone et al., 2014).

Critical race theory posits that historical and legal factors contribute to a culture of White supremacy and White privilege; this in turn leads to negative outcomes for people of color (Leonardo, 2004). These historical and cultural factors may help to account for differences in attrition and turnover; however, there is little empirical evidence regarding how rates of burnout may or may not differ across people of diverse races and ethnicities. This dearth in the literature may be partially explained by the relatively low proportions of teachers of color in general (Villegas et al., 2012). Regardless of the reason for the lack of research on burnout in teachers of color, this area represents a significant gap in the literature.

Environmental Factors

There is evidence that burnout is related to factors beyond individual differences. As mentioned previously, research examining how job resources and job demands impact burnout is prominent and suggests greater job demands are associated with greater burnout symptoms, whereas job resources may serve as a buffer for burnout (Alarcon, 2011). Unfortunately, the use of self-report data to measure "perceptions" of these constructs (e.g. perceptions of workload, perceptions of administrator support) confounds these factors with individual differences. Studies examining related constructs such as teacher turnover and teacher shortages, in

conjunction with research on burnout, help us to understand what environmental factors may relate to burnout.

Setting

Turnover also appears to relate to several external factors, such as geographical location, and type of setting (rural vs. urban vs. suburban). In the United States, the South has higher turnover compared to other regions, including the Midwest, Northeast, and West (Sutcher et al., 2016). Turnover is also higher in urban settings for most regions, compared to suburban or rural districts (Sutcher et al., 2016). As might be expected, areas with higher salaries are associated with lower attrition (Sutcher et al., 2016).

Working Conditions

A variety of working conditions have been found to predict teacher turnover and burnout symptoms (Grayson & Alvarez, 2008; Kokkinos, 2007; Loeb et al., 2005). Student characteristics, large class sizes, facilities problems, and lack of books, are associated teacher turnover (Loeb et al., 2005). Likewise, teacher-parent/community relation problems appear to be associated with increased Emotional Exhaustion, instructional management practices appear to be associated to Personal Accomplishment, and teacher relationships with students and administrators appears to be related to Cynicism/Depersonalization (Grayson & Alvarez, 2008). In fact, high-poverty schools tend to have higher rates of teacher turnover, likely due to poorer working conditions in these settings (Sutcher et al., 2016). Recent research has also examined the role of teacher-student relationships in predicting Personal Accomplishment and Emotional Exhaustion. Corbin et al., (2019) found that teacher-reported close relationships with students were associated with higher levels of Personal Accomplishment and self-reported conflictual relationships with students were associated with increased Emotional Exhaustion.

Certification and Training

Some research has found that alternatively certified teachers have higher rates of turnover compared to traditionally certified teachers (Robertson & Singleton, 2010; Sutcher et al., 2016). Lower standards for alternative certification programs may in part explain differences in attrition (Sutcher et al., 2016). While research on levels of burnout between alternatively and traditionally certified teachers is lacking, a recent dissertation examining Texas Special Education Certified teachers (n =210; 100 of whom obtained their certification from an alternative program) found that, contrary to their hypothesis, burnout did not predict method of certification (Casteel, 2018). In addition, recent research suggests that teaching students with disabilities is associated with turnover for general education certified teachers; however, teaching students with disabilities is not associated with turnover for special education certified teachers, suggesting general education teachers may need more support and training for working with students with disabilities (Gilmour & Wehby, 2019). Overall, it is unclear how certification may impact burnout symptoms.

Interventions

Currently, one other meta-analytic study has examined multiple types of interventions to treat teacher burnout (Iancu et al., 2018). Iancu et al. (2018) found that the summary effect size for Total Burnout Symptoms was small and statistically significant (d = 0.18, SE = 0.05, Z = 3.26, $p < .001$, k = 23). Similarly, Iancu et al. (2018) found that the effect sizes for Emotional Exhaustion and Personal Accomplishment were small and statistically significant, and the summary effect size for Cynicism/Depersonalization was “almost null” (d = 0.03, SE = 0.06, Z = 0.53, $p > .05$, k = 11).

Further, an analysis of potential moderators suggested that mindfulness interventions had significant effects on Emotional Exhaustion (d = 0.31; 95% CI 0.08, 0.54; p = <.01) and Personal Accomplishment; cognitive behavioral interventions had statistically significant effects on Emotional Exhaustion (d = 0.20; 95% CI −0.00, 0.41; p = <.05) only (Iancu et al., 2018). None of the intervention approaches appeared to produce statistically significant effects for Cynicism/Depersonalization (Iancu et al., 2018). Iancu et al. (2018) also found that effect sizes were smaller for primary and middle school teachers and effect sizes were smallest for interventions that lasted less than one month.

Interestingly, other meta-analyses of burnout interventions for other populations have found some consistent and some disparate results. One meta-analytic study found that both individual-focused interventions, such as mindfulness interventions, and organizational strategies were equally effective in preventing physician burnout (West et al., 2016).

Mindfulness

Mindfulness programs are the most studied interventions for preventing teacher burnout (Iancu et al., 2018; Zarate et al., 2019). Mindfulness programs vary between studies and sometimes include multiple components. Cultivating Awareness and Resilience in Education incorporates emotion skills instruction, compassion practices, and mindfulness training (Jennings, 2011; Jennings & Greenberg, 2009). Some research uses or adapts mindfulness-based stress reduction (MBSR; Kabat-Zinn, 1990). The Inner Resilience Pilot Program was developed to combat burnout and includes aspects of mindfulness training along with other components, such as social-emotional learning interventions, “stress reduction days,” parent and staff workshops, and residential retreats for staff (Lantieri & Malkmus, 2011). Many mindfulness-

based interventions teach body awareness, coping skills, and incorporate yoga (Ancona & Mendelson, 2014).

Mindfulness interventions have a strong theoretical link to teacher burnout. Mindfulness interventions target individuals directly, and individual teacher level factors explain most of the variance in burnout symptoms, above and beyond school level factors (McCarthy et al., 2009; Ullrich et al., 2012). Mindfulness and burnout are both related to personality. Meta-analyses examining burnout and personality and mindfulness and personality have shown that mindfulness and all three facets of burnout are strongly related to neuroticism, and at least modestly related to conscientiousness and agreeableness (Giluk, 2009; Swider & Zimmerman, 2010).

A growing body of research suggests mindfulness interventions are effective in reducing stress, depression, anxiety, and improving self-reported mental and physical health in a variety of populations, including teachers (Cavanagh et al., 2014; Chiesa & Serretti, 2009; Zarate et al., 2019). Burnout symptoms are also related to stress, depression, and physical health (Alarcon, 2011; Armon et al., 2010; Bianchi et al., 2013; Hallman et al., 2003; Melamed et al., 2006; Yu et al., 2015). In fact, a recent exploratory structural equation modeling bifactor analytic study examining burnout, depression, and anxiety scales suggest that anxiety, depression, and Emotional Exhaustion all load onto the same nonspecific psychological distress (NPD) factor, and thus, interventions that have evidence for depression and anxiety have a strong theoretical link to burnout reduction (Schonfeld et al., 2019a).

Most relevant to the relationship between mindfulness and burnout is the existing intervention research on burnout specifically. As mentioned previously, two meta-analyses on teacher burnout interventions by Iancu et al. (2018) and Zarate et al. (2019) both examined

mindfulness interventions and found small effects. Iancu et al. (2018) found that mindfulness interventions were effective in reducing Emotional Exhaustion and Personal Accomplishment, specifically. Another meta-analysis found that mindfulness interventions have been shown to reduce psychological distress in working adults (Virgili, 2015). Further, mindfulness interventions have been studied for their impact directly on burnout in teachers in several primary studies (Ancona & Mendelson, 2014; Anderson et al., 1999; Flook et al., 2013; Frank et al., 2015; Harris et al., 2014; Jennings et al., 2013).

Other Interventions

Other interventions beyond mindfulness-based interventions have been studied for their ability to impact teacher burnout symptoms. These include social-emotional learning curricula (Castillo, Fernández-Berrocal, & Brackett, 2013; Castillo-Gualda et al., 2017; Domitrovitch et al., 2016), cognitive behavioral approaches (e.g., rational emotive behavior therapy, problem solving training; Anderson, 2000; Ebert et al., 2014), classroom/school behavior management approaches (e.g. GBG, PBISplus; Bradshaw et al., 2012; Breeman et al., 2016; Domitrovitch et al., 2016) stress management/burnout prevention workshops and peer collaboration programs (Cooley & Yovanov, 1996), professional development for interdisciplinary civic education (Barr et al., 2015), and Expressive Writing (Anopchand, 2000). Most of these interventions have at some evidence of impact on burnout symptoms; however, mindfulness interventions appear to have more research.

Tools for Measuring Burnout

A variety of different instruments have been developed and used to measure burnout; all of these have between two and four dimensions or subscales (See Chapter III for more details). Many of these tools conceptualize burnout differently. The Maslach Burnout Inventory (MBI;

Maslach & Jackson, 1981) is by far the most frequently used tool to measure burnout overall, and for teachers, the Maslach Burnout Inventory Educator Survey (MBI-ES; Maslach, Jackson, & Leiter, 1996) is most often used. For the purposes of this study, Maslach and Jackons's (1981) tripartite conceptualization of burnout was used, though studies using other measures of burnout were included, unless they measured a different construct than one of three dimensions in the tripartite model.

In addition to the MBI, other measures of burnout include The Teacher Burnout Scale (Seidman & Zager, 1986), Blasé's (1982) model, The Copenhagen Burnout Inventory (Kristensen & Borritz, 1999; Kristensen et al., 2005), the Oldenburg Burnout Inventory (OLBI; Demerouti et al., 2003; Halbesleben & Demerouti, 2005), The School Burnout Inventory (Salmela-Aro et al., 2009), the Shirom Melamed Burnout Questionnaire (Kushnir & Melamed, 1992; Lundgren-Nilsson et al., 2012; Melamed et al., 1992). Each of these measures has some evidence of yielding reliable and valid data.

Gaps in the Current Research

The research base on employee burnout is growing and several meta-analytic studies examining burnout already exist (Alarcon, 2011; Aloe et al., 2014; Brewer & Shapard, 2004; Edmonson & Thompson, 2000; Halbesleben, 2006; Iancu et al., 2018; Ng & Feldman, 2012; Purvanova & Muros, 2010; Reichl et al., 2014; Shin et al., 2014; Shoji et al., 2016; Swider & Zimmerman, 2010). Several of these meta-analyses examine burnout in teachers specifically (Aloe et al., 2014; Edmonson & Thompson, 2000; Iancu et al., 2018; Zarate et al., 2019) and only one has examined the effects of diverse interventions on preventing or treating teacher burnout specifically (Iancu et al., 2018).

The literature comparing interventions to prevent teacher burnout to comparison conditions began in the 1990's; since then, one meta-analysis has attempted to synthesize the results of these studies to examine overall effects and moderators (Iancu et al., 2018). Iancu et al. (2018) began its search with 513 records to review after duplicates were removed and found no additional records from sources other than databases. After exclusions, Iancu et al., included and examined a total of 23 studies. As with all research, Iancu et al. (2018) is not without its limitations. Additional meta-analyses that examine intervention effects, utilizing improved methods, are imperative to synthesize what is known about intervention effectiveness, as well as to answer questions additional questions such as: for whom are interventions most effective, and, in what situations interventions are most effective?

Currently, there exists limited research examining the impact of school wide positive behavioral support and social-emotional learning curricula on burnout (Domitrovitch et al., 2016). Further, there is no experimental research on mentoring interventions that examines burnout symptoms as an outcome, however, one study by Cooley and Yovanov (1996) does use structured peer collaboration without designated mentor-mentee roles. Much research on mentoring programs does not utilize control groups, and the experimental research that does exist examines retention or attrition outcomes, not teacher burnout symptoms (Ingersoll & Kralik, 2004).

CHAPTER III

METHOD

Research Design

The present study is a systematic meta-analytic study, meaning we have systematically included and excluded available evidence pertaining to the specific topic in order to answer our research questions. Meta-analyses are one approach to summarizing research findings. Systematic meta-analyses have an advantage over other methods of summarizing research for several reasons. A systematic approach of identifying studies, as is done in meta-analysis, has an advantage over narrative reviews due to the systematic search procedures used, which help to eliminate bias and make procedures replicable (Lipsey & Wilson, 2001). This study will follow the guidelines set by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009) to make the procedure explicit and replicable. In addition, meta-analyses have an advantage over other systematic reviews in that they can quantify the effect of an intervention in question. Conclusions from reviews without a quantitative approach may introduce bias from authors trying to interpret results without an objective weighting procedure. Meta-analysis allows researchers to estimate a summary effect size, or estimated mean effect size, of the intervention in question, using the available research, thus producing a more objective, replicable result than a traditional narrative review. Meta-analyses can be conducted with as few as two studies, though meta-analyses with smaller sample sizes may be of limited utility and must be interpreted with caution (Lipsey & Wilson, 2001).

Procedure

The present study followed the guidelines set forth by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009). The exact procedure is described below.

Protocol and Registration

This proposal was used as a guide for registering the present study protocol with PROSPERO (Center for Reviews and Dissemination). We described the rationale, hypothesis, and planned methods. The study protocol was made publicly available to assist in dissemination of the work.

Information Sources

To obtain relevant studies, the author searched the EBSCO database within ERIC, PsycINFO, and Academic Search Ultimate using keywords for the population, study design, and outcome. In addition to searching these databases, the author used Dissertation Abstracts to find relevant dissertations. The author examined the reference lists of included studies and relevant review articles to obtain additional studies.

Search

Search terms are listed in table 1. Search terms were connected with "OR" to expand the search using synonyms, and "AND" to limit the search to relevant articles. Asterisks were used to include multiple endings, such as "teachers," "teaching," "resilience," "resiliency," and so on.

Table 1 Search Terms

Population	"elementary school teachers" OR "middle school teachers" OR "secondary school teachers" OR "teacher*" OR "educator*"
	AND
Study Design	"experimental study" OR "quasiexperimental study" OR "randomized control" OR "quasiexperimental design" OR "program effectiveness" OR "experimental groups" OR "intervention" OR "mindfulness" OR "Social-emotional learning" OR "SEL" OR "relaxation training"
	AND
Outcome	"teacher burnout" OR "teacher morale" OR "Resilience (psychology)" OR "burnout" OR "morale" OR "mental health" OR "resilien*" OR "teacher retention" OR "teacher satisfaction"

Study Selection

The first author screened study titles and abstracts using Rayyan (Ouzzani et al., 2016), and the eligibility criteria. Studies that appeared to meet criteria were obtained in full text and the method and results sections were screened further to determine if the study met inclusion criteria (see figure 1).

Data Collection

The author used the coding sheet (see Appendix A) to retrieve relevant information on each study included. The dissertation chair, who had experience conducting a meta-analysis, was trained to code the studies as a second coder for 9 the 30 included studies (30%). The nine studies coded by the seconder coder were randomly selected using a random number generator.

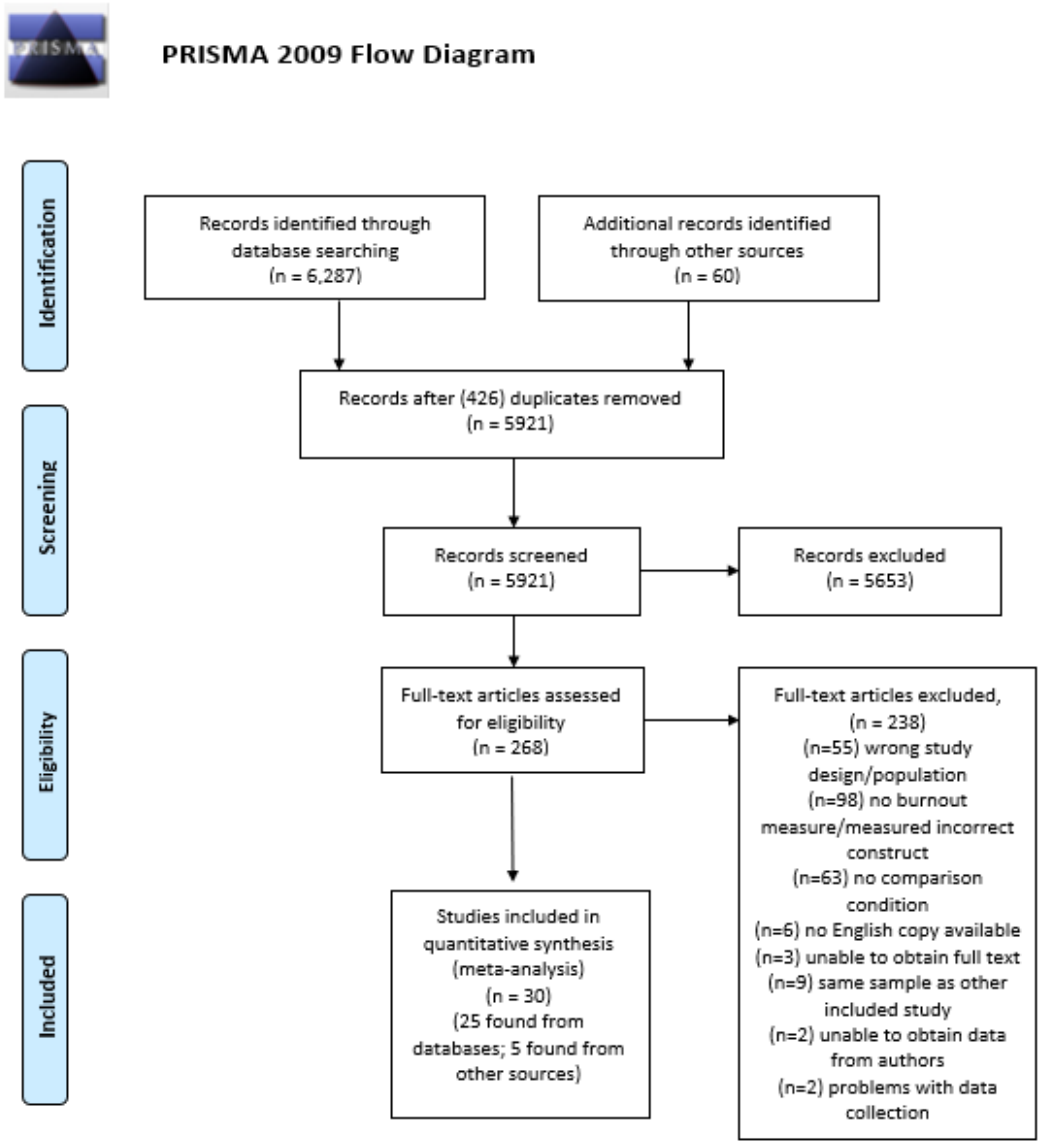

Figure 1 PRISMA Flow Chart (Moher et al., 2009)

Data Items

We used a standardized coding sheet to extract information on effect sizes and other statistics. Information on all moderators was extracted using this coding tool (see Appendix A). Interrater reliability for the coding tool was 90.34%. The coding sheet also included a section dedicated to measuring study quality (see Appendix A). This study quality indicator was developed by the present study authors and is based on information provided in the "Quality indicators for group experimental and quasi-experimental research in special education" developed by Gersten et al. (2005). Interrater reliability for the study quality indicator alone was initially 79.37%. Several items in the study quality indicator had particularly low inter-rater reliability. We deleted three of these items from the measure. The first author re-wrote two of these items to make the items more objective and simpler to code and then re-coded these two items. The interrater reliability for the remaining items that were not deleted or re-coded, was 89.63%. All Discrepancies were resolved by discussion as well as the first author re-examining included studies to decide how to code discrepant items. The first author determined the final coding for each item before beginning analyses.

Risk of Bias in Individual Studies

Risk of bias in individual studies was assessed in several ways (see Appendix A). We extracted data regarding type of control group (active vs. passive), how participants were assigned to condition (random, matching, or convenience), whether the study was peer reviewed, the type of report (conference paper, journal article, dissertation, masters' thesis, or other). We conducted meta-regressions to examine if type of control group (dummy coded) and use of random assignment (dummy coded) were categorical moderators.

Planned Analyses and Synthesis of Results

A random effects model was used to estimate individual study and summary effect sizes. A random effects model assumes study samples may be taken from different populations and is often considered the preferable model for conducting meta-analyses (Borenstein et al., 2011). In comparison to a fixed effect model, the random effects model may lead to less weight applied to large studies, a wider confidence interval about each coefficient, decreased power to detect statistical significance, and thus decreased risk of Type I error (Borenstein et al., 2009). We extracted study effect sizes, computed summary effect sizes for each facet of burnout, assessed heterogeneity across study effect sizes, and examined the potential effects of moderators at the study level.

Participants/Studies

Studies identified based on the search criteria described in the procedure are the units of analysis. Thirty studies are included in the present study (k = 30), and the total number of participants included across the included studies was over three thousand (n = 3024). The PRISMA flow chart for excluded studies is presented in Figure 1 (Moher et al., 2009). The present study includes only studies evaluating the efficacy of an intervention for its effects on Emotional Exhaustion, Personal Accomplishment, or Cynicism/Depersonalization, or another burnout score, conceptualized as "total burnout" (Maslach & Johnson, 1981) or "work-related burnout" (as used in the CBI; Kristensen & Borritz, 1999; Kristensen et al., 2005). Type of intervention was not used as an exclusion criterion. Research on a variety of intervention types were included as a part of this study, including, mindfulness (with or without meditation or yoga); behavior management interventions; social emotional learning; bibliotherapy; psychoeducation; cognitive behavioral therapy groups; and internet-based stress management

interventions. Only experimental or quasi-experimental studies were included; that is, included studies have both a treatment and a control group and collected both pre-intervention and post-intervention testing on the same participants.

The participants of the included studies were teachers of elementary, secondary, or high school. Studies examining preschool teachers, university faculty, or teachers of other adult classes were excluded from analyses. Only studies written or translated into English were included for practical reasons. Review articles, case studies, and studies only examining related constructs, such as depression and anxiety, were excluded. As stated previously, some measure of burnout was required for inclusion.

Both published and unpublished works were included to limit publication bias. Single case research and single group studies were excluded to ensure effect sizes are consistent and based on experimental or quasi-experimental research. Studies conducted prior to 1990 or after 2017 were omitted, as intervention research on burnout did not begin until the empirical phase within the burnout literature around 1990, and the study search phase of the meta-analysis ended in 2017. Studies that reported problems with data collection (Wolf et al., 2015) such that post-test was conducted on participants who were not included in the pre-test, also were excluded. Finally, one study was excluded after data extraction (Cheon et al., 2014); this study's results were initially included in Emotional Exhaustion results; however, this study produced a higher effect size compared to other studies. After reviewing the information provided in the study to ensure there was not an error in data entry, we determined we would exclude this study as it used the Athlete Burnout Questionnaire (ABQ; Raedeke & Smith, 2001) to measure physical and emotional exhaustion, which appears to be a different construct than Emotional Exhaustion traditionally measured by other burnout measures such as the Maslach Burnout Inventory (MBI;

Maslach & Jackson, 1981; MBI-ES, Maslach et al., 1996). Thus, results are presented in the present study exclude Cheon et al. (2014).

Instruments

Maslach Burnout Inventories

While we did not specify the instrument used, all but three included studies used the Maslach Burnout Inventory (MBI; Maslach & Johnson, 1981) or the Maslach Burnout Inventory- Educator Survey (MBI-ES; Maslach et al., 1996) as an outcome measure.

The MBI-ES has been shown to yield reliable and valid data. Results of exploratory and confirmatory factor analysis support the three-factor model, and Cronbach's alpha for the entire scale was found to be .74 for a sample of 771 teachers in Cyprus (Kokkinos, 2006). Both the MBI and MBI-ES have 22 items and three dimensions: Emotional Exhaustion, Cynicism/Depersonalization, and Personal Accomplishment. In research, a shortened, 12 item version of the MBI is sometimes used, and other times, researchers select only one dimension of burnout to assess, typically Emotional Exhaustion. While the MBI and MBI-ES have evidence of reliability and an adequate factor structure (Maslach & Jackson, 1981; Maslach et al., 1996), the Maslach Inventories are often criticized for the way items are worded; each dimension uses only positively or negatively worded items (Demerouti et al., 2001). Wording items in this way presents a large problem from a psychometric standpoint as it can lead to artificial factor solutions (Demerouti et al., 2001).

Other Burnout Measures

Two included studies used the Copenhagen Burnout Inventory (CBI; Kristensen & Borritz, 1999); both studies were conducted by Johnson and Naidoo (2013, 2017). The CBI has

19 items across three dimensions (personal burnout, work-related burnout and client-related burnout) and has some evidence for yielding reliable and valid data (Kristensen et al., 2005).

Cheon et al., (2014) used the Athlete Burnout Questionnaire (ABQ; Raedeke & Smith, 2001) to measure physical/emotional exhaustion. These scores were coded and analyzed along with other measures of Emotional Exhaustion; however, this study was ultimately excluded due to measuring a different construct than the specific construct of interest.

Study Effect Sizes

To examine post-intervention and follow-up differences between intervention and control groups, we computed standardized mean difference effect sizes, Hedges *g*, for each study and each facet of burnout (Lipsey & Wilson, 2001). We then computed a 95% confidence interval for each study effect size.

The equation for calculating a standardized mean difference effect size is presented below (Lipsey & Wilson, 2001). In this equation, ES is the standardized mean difference effect size (unadjusted for bias), Mg1 is the mean for group one (i.e. either treatment or control group), Mg2 is the mean for group two, and sp is the pooled standard deviation for both groups (equation 1).

$$ES = \frac{Mg1 - Mg2}{sp} \quad (1)$$

The equation for the pooled standard deviation (sp) is presented below (Lipsey & Wilson, 2001). For this equation, n1 is the sample size for group 1, n2 is the sample size for group 2, sd1 is the standard deviation for group 1 and sd2 is the standard deviation for group 2 (equation 2)

$$sp = \sqrt{(\frac{(n1-1)sd1\text{^}2+(n2-1)sd2\text{^}2}{(n1-1)+(n2-1)})} \quad (2)$$

Hedges g is adjusted for bias due to small sample sizes (Hedges, 1981; Lipsey & Wilson, 2001). The equation for calculating Hedges g from the biased standardized mean effect size is presented below, where ES is the unadjusted standardized mean effect size and N is the total study sample size (equation 3).

$$g = ES\ [1 - \left[\frac{3}{4N-9}\right]] \quad (3)$$

For studies that that failed to report means and standard deviations, we contacted authors via email to obtain this information. If authors did not respond after three attempts, we estimated Hedges g based on the effect size statistic provided. If no effect size statistic was provided, we excluded the study. Barr et al., (2015) did not report standard deviations for their post-test results; however, they did report Glass's Δ (delta). Glass's Δ was used to estimate Hedges g using equation 2, where ES was Glass's Δ.

Summary Effect Sizes

The summary effect size statistics used for this study are standardized mean difference effect size estimates, specifically, Hedges g. Summary effect sizes were computed for the three scales within the MBI-ES, at post-intervention, calculated as Hedges g. The standardized mean difference effect size for each study, and for each facet of burnout (Emotional Exhaustion, Cynicism/Depersonalization, and Personal Accomplishment) are presented in the Results section, along with 95% confidence intervals, the summary effect size statistics, and the standard errors of the summary effect size statistics. Hedges g is used as the standardized mean difference score for each facet of burnout to correct for upward bias due to small sample sizes (Hedges, 1981; Lipsey & Wilson, 2001).

Although Hedge's *g* was utilized as the effect size, Cohen's conventions often are used to conservatively interpret Hedges' *g* (Borenstein et al., 2011). Cohen's *d* and Hedge's *g* are similar in many respects; both are standardized mean difference effect sizes; however, Hedges g includes a correction for upward bias as shown in equation 3. For the purposes of this study, effect sizes were categorized as small (0.2) medium (0.5) or large (0.8), based on Cohen's conventions (Cohen, 1988). Despite this categorization, it is important to understand that the conventions are arbitrary and should not be interpreted rigidly; rather effect sizes should be interpreted within the context of other effects in the relevant literature (Thompson, 2002).

Durlak (2009) offers three guidelines for evaluating effect sizes in context. The first guideline involves considering the source (i.e. quality) of the research that yields an effect. The second guideline recommends making comparisons across similar research conditions (e.g., using the same outcome measure, measuring the same construct). The final guideline recommends researchers consider the clinical or practical significance of the effect size (Durlak, 2009). This meta-analysis attempts to address both the first and second guidelines by evaluating the quality of the included studies and by comparing our obtained summary effect sizes with other effect sizes obtained for reducing burnout in teacher populations. In this case, we will compare our obtained summary effect sizes to those summary effect sizes obtained in another recent meta-analysis conducted by Iancu et al., (2018). Following the third guideline presented by Durlak presents a greater challenge; it is not always feasible to establish what size effects have practical significance.

Each summary effect size statistic for each facet of burnout (Emotional Exhaustion, Cynicism/Depersonalization, and Personal Accomplishment) was computed by summing the product of study effect sizes (also computed as Hedges *g*) and their weights, and then dividing

this sum by the sum of the weights. Values of the weights for each study are determined by the type of model chosen, and in this case, a random effects model was used to compute the summary effect size statistic. Individual study effect sizes were computed in Microsoft Excel. The statistical program R was used to carry out all further analyses (R Core Team, 2019).

The weights for each study were computed using a random effects model, assuming sources of variability are randomly distributed (Lipsey & Wilson, 2001). If the effect size distribution for all the studies included is, in fact, homogenous, then results would be the same as in a fixed effect model (Langan et al., 2019; Lipsey & Wilson, 2001).

The random-effects model suggests that individual study effect size estimates vary from the true mean effect size estimate for two reasons: inherent heterogeneity or variation between studies and within study error. To help understand this model, we present two equations (equation 4 and 5) below.

For equation 4, the true effect size for an individual study is represented by θi; θei is the estimated effect size for that same study, and εi is the within-study error. That is, the true effect size of a study, plus the within-study error, equals the estimated effect size. All individual study effect size estimates differ from the true individual study effect size due to error.

$$\theta ei = \theta i + \varepsilon i \qquad (4)$$

For equation 5, the true effect size, θi, of an individual study equals the mean effect size across studies, θm, plus between-study heterogeneity, δ (Borenstein et al., 2011; Langan et al.,

2019) Taken together, these equations (equations 4 and 5) demonstrate that, for a random effects model, individual study effect size estimates are assumed to contain within-study error, and additionally, there is assumed to be variation in the true study effect sizes across studies (i.e. between-study heterogeneity). In the results, we present estimates of the heterogeneity variance parameter is represented by (τ^2). Of note, τ^2 estimates provided in the Results section are not the true heterogeneity variance parameters as these population parameters are unknown; rather, they are statistics—estimates based on our models.

$$\theta i = \theta m + \delta \quad (5)$$

We used the Hartung-Knapp-Sidik-Jonkman (HKSJ) method of computing the mean effect sizes, confidence intervals, and the heterogeneity variance estimate (τ^2) as this method tends to produce estimated summary effect sizes and confidence intervals that are more robust to fluctuations in the heterogeneity variance estimate (τ^2) than some other methods (Langan et al., 2019). For example, the HKSJ method has been shown to outperform the DerSimonian-Laird (DL) method in many cases (IntHout et al., 2014; Langan et al., 2019). Overall, the HKSJ method tends to lead to wider confidence intervals (i.e. more conservative results; IntHout et al., 2014).

We set alpha at .05, meaning *p*-values less than .05 would be considered statistically significant. While statistical significance is often conflated with the importance of a finding, our focus in interpreting effect sizes will be on the size of the effect, not on whether results are statistically significant according to our alpha level.

Heterogeneity in Effect Sizes

Heterogeneity across studies was examined in several ways. First, we examined overall heterogeneity for each outcome by calculating the *Q* statistics, I^2 statistics, τ^2, and interpreting forest plots. We reported summary effect sizes along with 95% confidence intervals and estimates of τ^2, which we computed using the HKSJ method within the metagen package in R (Möbius, 2014; R Core Team, 2013). The I^2 statistics were also computed using the HKSJ method for estimating τ^2; the metagen package in R uses derives the I^2 statistic from τ^{2-} as opposed to deriving this statistic from the *Q* statistic. Moderator analyses also were conducted to explore potential causes of systematic variance using the metafor package for R (Viechtbauer, 2010).

We used a different method, Restricted Maximum likelihood (REML), to estimate τ^2 when conducting the meta-regressions to examine potential moderators, thus, heterogeneity variance estimates sometimes differed depending on the estimation method for τ^2. We used both REML and HKSJ in the present study because we found that, during preliminary analyses, HKSJ yielded higher heterogeneity estimates and we wanted to use this more conservative approach to examine the mean effect sizes and confidence intervals. We chose to use REML as the method for calculating heterogeneity estimates when examining moderators using meta-regressions in an attempt to avoid over-estimating the heterogeneity variance accounted for by the model.

The *Q* statistic is the weighted sum of square deviations from the mean effect size. It is computed by multiplying each study's squared deviation from the mean by its inverse variance weight, and then summing those values (Borenstein et al., 2011). The *Q* statistic is not a particularly intuitive measure to interpret; it is a sum and depends strongly on the number of studies. A statistically significant *p* value for the Q statistic does provide evidence that true

effects appear to vary across studies; however, a non-statistically significant p-value does not mean that heterogeneity is low; it may simply be that a small number of studies are included or that within-study variance is large (Borenstein et al., 2011)

It I^2 statistic helps to answer the question: What proportion of the observed variance reflects real differences in effect sizes?" (Borenstein et al., 2011, p.117). I^2 has a range of 0-100%; values of approximately 25% are considered low, whereas 50% might be considered moderate and 75% may be considered high. I^2 values of 100% indicate only that most of the observed variance is likely to be true variance rather than spurious variance; high values do not mean that heterogeneity is high nor do low values mean that between study variation is low (Borenstein et al., 2011).

Examining Moderators

We used meta-regression to examine categorical moderators (dummy coded) and continuous moderators. Meta-regression, like multiple regression, assesses the relations between one or more predictors and a dependent variable. For meta-regression, the "predictors" are moderators at the study level and the dependent variable is the effect size for the study (Borenstein et al., 2009). Meta-regression is not recommended when the number of studies is small (Borenstein et al., 2009). We examined the following continuous variables as moderators of intervention effectiveness: length of direct intervention, total time in minutes of direct contact, weeks of homework assigned, study quality, mean years of teaching experience, mean age of participants, percent of females in each study, percent of males in each study, and percent of White participants in each study. Categorical moderators included type of intervention and socio-economic status, as well as type of control group and use of random assignment.

We used *Q*-tests and goodness of fit tests to test our models. The *Q* statistic is the weighted sum of squares and reflects the total variability of studies. For a fixed effect model, the *Q*-test partitions *Q* into its component parts, Q_{resid} (Q_R) and Q_{model} (Q_M) such that Q_R and Q_M are additive; however, for a random effects model, the weights assigned for each study incorporate between study variance; thus, the variance components are not additive for the random effects model (Borenstein et al., 2009). The random effects model assumes that for any value of a moderator, there is a distribution of true slope coefficients and the true coefficient depends on the subgroup of the population; the slope coefficient (*B*) found for each moderator is assumed to be the mean, not the "true" coefficient (Borenstein et al., 2009).

For each moderator examined we present the results of the *Q* test and the goodness of fit test. For the *Q* test, we examine if the p-value that corresponds to Q_M is statistically significant at alpha = .05. For the goodness of fit test, we examine Q_R, and its corresponding p value, as well as τ^2, to examine heterogeneity not explained by the model. In this context, τ^2 refers to the estimated residual heterogeneity variance not explained by the model. We also examine I^2,which in the context of random effects meta-regression, refers to the proportion of the unexplained variance that is likely true variance as opposed to error. We also examine R^2, the proportion of the variance that is explained by the model. The following sections describe the plan for how each hypothesis was to be tested specifically.

Hypothesis 1

It was hypothesized that overall, interventions would produce at least a small effect in reducing burnout, as defined using Cohen's conventions (1988). This hypothesis was tested by calculating the summary effect size statistic for all included studies, using a random effects model. The effect size used was Hedges *g*, a standardized mean difference effect size. Hedges *g*, as a standardized mean effect size, quantifies the difference in means between two groups post-

intervention. Further, Hedges *g* is distinct from other standardized mean effect sizes in that it includes an adjustment for bias caused by small samples. Thus, Hedges *g* was calculated or estimated based on the information provided in each study. For studies that examined more than one intervention group, only one group was randomly selected to be included in analyses so as to maintain independence of control groups; this was the case for two included studies (Dicke et al., 2015; Domitrovitch et al.,2016). The weighted average effect size, computed using a random effects model, is the summary effect size. We computed the 95% confidence interval about this mean. If the summary effect size is equal to or greater than .20 then hypothesis 1 was supported (Cohen, 1988).

Hypothesis 2

We hypothesized that only trivial differences will remain at follow-up, that is, an average standardized mean difference effect size smaller than .20 (Cohen, 1988). Follow-up statistics were extracted from each study, when available (see Appendix A). Hedges *g* was computed or estimated for each study's follow-up means. The summary effect size for follow-up data was calculated as the weighted average of the effect sizes for follow-up data and was calculated using a random effects model. If the summary effect size for follow-up is less than .20 then hypothesis 2 was supported (Cohen, 1988).

Hypothesis 3

We hypothesized that mindfulness interventions targeting teachers will have larger effects than classroom management or other behavioral interventions. Studies was coded for the type of intervention used (see Appendix A). Interventions was compared using meta-regression if sufficient studies are available. If less than ten studies were included, we planned to use the analogue to the ANOVA test, which allows comparisons amongst categorical or nominal

variables. We planned to test this hypothesis only if sufficient studies could be categorized as the same type of intervention; generally speaking, there must be at least two studies per intervention for interventions to be compared. We tested this hypothesis by examining Q_{model} (Q_M) to determine if it was statistically significant (alpha = .05) in the expected direction. We also examined $Q_{Residual}$(Q_R) to test the goodness of fit of the model.

Hypothesis 4

We hypothesized that several factors, including length of intervention, use of homework, work experience, age, sex, and race may moderate effectiveness. See hypotheses 4a through 4h.

Hypothesis 4a

We hypothesized that studies with interventions implemented over longer periods of time would produce larger effect sizes for each facet of burnout. We coded intervention length in several ways for each study (see Appendix A). We entered intervention length and minutes of direct contact as continuous variables conducted meta-regressions for both length of intervention and minutes of direct contact, independently. This hypothesis was tested by examining if $Q_{\text{model's}}$ (Q_M) were statistically significant (alpha = .05) in the expected directions. We also examined $Q_{Residual}$(Q_R) to test the goodness of fit of the model.

Hypothesis 4b

We hypothesized that interventions that incorporated homework would produce larger effects for each facet of burnout. Homework assigned and completed was coded for each study (see Appendix A). We conducted meta-regressions for the facets of burnout which included sufficient data for a meta-regression, with weeks of homework as a continuous moderator. This hypothesis was tested by examining if Q_{model} (Q_M) was statistically significant (alpha = .05) in the expected direction. We also examined $Q_{Residual}$(Q_R) to test the goodness of fit of the model.

Hypothesis 4c

We hypothesized that samples with a lower mean for years of work experience would see a greater reduction in burnout symptoms for all three facets of burnout. We coded mean work experience for each study (see Appendix A), entered mean work experience as a continuous variable, and conducted meta-regressions for the facets of burnout which included sufficient data for a meta-regression. This hypothesis was tested by examining if Q_{model} (Q_M) was statistically significant (alpha = .05) in the expected direction. We also examined $Q_{Residual}$ (Q_R) to test the goodness of fit of the model.

Hypothesis 4d.

We hypothesized that samples with a lower mean age would see greater reductions in Emotional Exhaustion and Cynicism/Depersonalization compared to samples with lower proportions of younger teachers. We coded mean teacher age for each study (see Appendix A) and conducted meta-regressions for the facets of burnout which included sufficient data for a meta-regression with mean teacher age as continuous variable. This hypothesis was tested by examining if Q_{model} (Q_M) was statistically significant (alpha = .05) in the expected direction. We also examined $Q_{Residual}$ (Q_R) to test the goodness of fit of the model.

Hypothesis 4e

We hypothesized that samples with higher proportions of female teachers would see a greater reduction in Emotional Exhaustion compared to samples with lower proportions of female teachers. We coded the percentage of female teachers for each study, if data were available (see Appendix A). We conducted a meta-regression on Emotional Exhaustion with the percent of female teachers as a continuous moderator. This hypothesis was tested by examining

if Q_{model} (Q_M) was statistically significant (alpha = .05) in the expected direction. We also examined $Q_{Residual}$ (Q_R) to test the goodness of fit of the model.

Hypothesis 4f

We hypothesized that samples with higher proportions of male teachers would see a greater reduction in Cynicism/Depersonalization compared to samples with lower proportions of female teachers. As shown in Appendix A, we coded the percentage of female, male, and non-binary teachers for each study. We planned to conduct a meta-regression with the percentage of males as a continuous moderator for Cynicism/Depersonalization. We planned to test this hypothesis by examining if Q_{model} (Q_M) was statistically significant (alpha = .05) in the expected direction and by examining $Q_{Residual}$ (Q_R) to test the goodness of fit of the model; however, we ultimately did not test this hypothesis due to missing data.

Hypothesis 4g

We hypothesized that samples with higher proportions of White teachers would show greater reduction in burnout symptoms across domains compared to samples with lower proportions of White teachers. We coded the percentage of White/Caucasian for each study (see Appendix A). The proportion of White teachers was entered as a continuous variable and we conducted meta-regressions for the facets of burnout which included sufficient data for a meta-regression. We tested this hypothesis by examining if Q_{model} (Q_M) was statistically significant (alpha = .05) in the expected direction. We also examined $Q_{Residual}$ (Q_R) to test the goodness of fit of the model.

Hypothesis 4h

We hypothesized that samples conducted in schools with lower proportions of students with free and reduced lunch will show greater reductions in burnout symptoms across domains.

We coded the proportion of students receiving free and reduced lunch (see Appendix A). We planned to enter the proportion of students receiving free and reduced lunch as a continuous variable; however, a meta-regression was not conducted due to insufficient data. Instead, socio-economic status (SES) was coded based on whether the studies reported that the student population was majority low SES or majority high/middle SES and an analogue to ANOVA was used to compare the effect sizes from low SES and high/middle SES populations. We planned to test this hypothesis by examining if Q_{Between} (Q_B) was statistically significant (alpha = .05) in the expected direction and examined $Q_{Residual}$(Q_R) to test the goodness of fit of the model; however, we ultimately did not test this hypothesis due to missing data.

Risk of Bias Across Studies

We conducted meta-regressions to examine study quality, as measured by our study quality indicator, as a continuous moderator for each facet of burnout and Total Burnout symptoms. Further, assessment of bias was conducted by examining forest plots for asymmetry. We created a forest plot for each set of summary effect sizes, on each dimension of burnout.

CHAPTER IV

RESULTS

The planned analyses describe in Chapter III were completed. For the 30% of studies that were coded by two coders, overall interrater reliability was acceptable at 90.34%. All effect sizes were computed such that a positive value indicates improvement for the treatment group; that is, lower levels of in Emotional Exhaustion, Cynicism/Depersonalization, and Total Burnout symptoms in the treatment group are represented by positive effect sizes, and higher levels of Personal Accomplishment for the treatment group are represented by positive effect sizes. Initial considerations were for heterogeneity. Heterogeneity across studies was examined by calculating the Q statistics, I^2 statistics, and interpreting forest plots. Moderator analyses also were conducted to explore potential associations with any systematic variance. Study quality was examined as a continuous moderator following the testing of hypotheses. The following sections present the results.

Characteristics of Included Studies

Table 2 presents information on the characteristics of included studies. Appendix B presents a supplementary table of demographic characteristics for included studies.

Table 2 Study Characteristics

Citation	Type of Research Record	Peer Review Status	Burnout measure	Sample Size	Intervention	Weeks Direct Contact (including latency)	Weeks of Homework
Ancona & Mendelson, 2014	Journal Article	Peer Reviewed	MBI-ES	n = 43	6-session yoga and mindfulness intervention for teachers. The intervention was developed by the Holistic Life Foundation (HLF)	3 weeks	3 weeks
Anderson et al., 1999	Journal Article	Peer Reviewed	MBI-ES	n = 91	Standard Meditation	5 weeks	5 weeks meditation
Anderson, 2000	Dissertation	Not Peer Reviewed	MBI-ES	n = 103	The Rational Emotive Behavior Therapy Bibliotherapy Intervention	0; No direct contact	Not reported
Anopchand et al., 2000	Dissertation	Not Peer Reviewed	MBI-ES	n = 66	Expressive Writing	2 weeks	0 weeks

Table 2 Continued

Citation	Type of Research Record	Peer Review Status	Burnout measure	Sample Size	Intervention	Weeks Direct Contact (including latency)	Weeks of Homework
Barr et al., 2015	Journal Article	Peer Reviewed	MBI-ES	n = 113	Facing History and Ourselves	Not reported	Not reported
Bradshaw et al., 2012	Journal Article	Peer Reviewed	4 items from MBI Emotional Exhaustion only	n = 729	PBISplus	72 weeks	Not reported
Breeman et al., 2016	Journal Article	Peer Reviewed	Dutch adaptation of MBI (UBOS-L; Schaufeli & Van Dierendonck, 2000)	n = 58	Good Behavior Game	40.5 weeks	Not reported
Castillo et al., 2013	Journal Article	Peer Reviewed	MBI-ES Spanish version	n = 47	RULER Approach to Social and Emotional Learning	26 weeks	Not reported
Castillo-Gualda, et al., 2017	Journal Article	Peer Reviewed	MBI-ES Spanish version	n = 54	RULER Approach to Social and Emotional Learning	13.5 weeks	Not reported
Cheek et al., 2003	Journal Article	Peer Reviewed	MBI-ES	n = 51	Cognitive Behavioral/music therapy (active control group received CBT)	6 weeks	5 weeks (find music for group)

Table 2 Continued

Citation	Type of Research Record	Peer Review Status	Burnout measure	Sample Size	Intervention	Weeks Direct Contact (including latency)	Weeks of Homework
Cooley, 1995	Final report for federally funded project.	Not Peer Reviewed	MBI	n = 67	(Two parts) Stress Management/Burnout Prevention Workshops and Peer Collaboration Program	10 weeks	Not reported
de Carvalho et al., 2017	Journal Article	Peer Reviewed	MBI-ES Portuguese version (Marques Pinto et al. 2005).	n = 20	MindUp	Not reported	15 weeks
Dicke et al., 2015	Journal Article	Peer Reviewed	Short, German version of the MBI-ES ; Emotional Exhaustion scale (Enzmann & Kleiber, 1989).	n = 61	Stress Management Training (Classroom management intervention was not coded)*	12 weeks	Not reported

Table 2 Continued

Citation	Type of Research Record	Peer Review Status	Burnout measure	Sample Size	Intervention	Weeks Direct Contact (including latency)	Weeks of Homework
Domitrovich et al., 2016	Journal Article	Peer Reviewed	Two subscales from the MBI	n = 249	PAX Good behavior game GBG (PATHS to PAX was not coded)*	31 weeks	Not reported
Ebert et al., 2014	Journal Article	Peer Reviewed	MBI	n = 133	Internet-based problem-solving training (iPST)	0; No direct contact	5 weeks
Elder et al., 2014	Journal Article	Peer Reviewed	MBI-ES	n = 36	Transcendental Meditation	18 weeks	18
Emery, 2011	Dissertation	Not Peer Reviewed	MBI-ES	n = 35	Acceptance and Commitment Therapy	1 week	0 weeks
Erdman, 2014	Dissertation	Not Peer Reviewed	MBI-ES	n = 42	Online Improving Your Social Environment Teacher Burnout Syndrome Amelioration	0; No direct contact	6 weeks

Table 2 Continued

Citation	Type of Research Record	Peer Review Status	Burnout measure	Sample Size	Intervention	Weeks Direct Contact (including latency)	Weeks of Homework
Flook et al., 2013	Journal Article	Peer Reviewed	MBI-ES	n = 18	Mindfulness Based Stress Reduction adapted for elementary school teachers	8 weeks	8 weeks
Frank et al., 2015	Journal Article	Peer Reviewed	MBI-ES	n = 36	Mindfulness Based Stress Reduction	8 weeks	8 weeks
Harris, 2014	Dissertation	Not Peer Reviewed	MBI-ES	n = 63	CALM (The Comprehensive Approach to Learning Mindfulness)	17 weeks	17 weeks
Jennings et al., 2013	Journal Article	Peer Reviewed	MBI-ES	n = 50	Cultivating Awareness and Resilience in Education (CARE)	8.5 weeks	Not reported
Jennings et al., 2017	Journal Article	Peer Reviewed	MBI-ES Emotional Exhaustion only	n = 200	Cultivating Awareness and Resilience in Education (CARE for Teachers)	18 weeks	Not reported

Table 2 Continued

Citation	Type of Research Record	Peer Review Status	Burnout measure	Sample Size	Intervention	Weeks Direct Contact (including latency)	Weeks of Homework
Johnson & Naidoo, 2013	Journal Article	Peer Reviewed	Copenhagen Burnout Inventory (CBI; Kristensen & Borritz, 1999)	n = 54	Capacitar transpersonal psychology workshops	Not reported	Not reported
Johnson & Naidoo, 2017	Journal Article	Peer Reviewed	Copenhagen Burnout Inventory (CBI; Kristensen & Borritz, 1999)	n = 36	Transpersonal psychology	10 weeks	Not reported
Lantieri & Malkmus, 2011	Research Report for Metis Associates, an independent research and evaluation firm	Not Peer Reviewed	MBI-ES	n = 57	Inner Resilience Pilot Program (Nurturing the Inner Life meetings + yoga, retreat, and student curriculum)	Not reported	Not reported
Porter, 2000	Dissertation	Not Peer Reviewed	MBI-ES	n = 92	psychoeducational groups	6	0
Roeser et al., 2013	Journal Article	Peer Reviewed	MBI	n = 113	Mindfulness Training	8	8
Siu et al., 2014	Journal Article	Peer Reviewed	MBI	n = 98	"stress management"	1	Not reported

Table 2 Continued

Citation	Type of Research Record	Peer Review Status	Burnout measure	Sample Size	Intervention	Weeks Direct Contact (including latency)	Weeks of Homework
Unterbrink et al., 2012	Journal Article	Peer Reviewed	German version of MBI-D	n = 209	manual-based psychological group program	45	Not reported

Heterogeneity Analyses

Heterogeneity across studies was examined by calculating the *Q* statistics, I^2 statistics, the estimate of the between-studies variance component (τ^2) and interpreting forest plots. The *Q* statistics, I^2 statistics, and τ^2 estimates were computed using the HKSJ method within the metagen package in R (R Core Team, 2013; Möbius, 2014). Moderator analyses were also conducted to explore potential causes of systematic variance using the rma function in the metafor package for R (Viechtbauer, 2010).

Heterogeneity of Post-Intervention Results

Results are presented first for effect sizes at post-intervention. Table 3 provides an overview of relevant statistics for examining heterogeneity at post-intervention.

Table 3 Effect Sizes at Post-Intervention

Scale	*k*	Hedges *g*	95% CI		*t*	τ^2	*Q*	I^2
Emotional Exhaustion	25	0.16	0.09	0.23	4.70***	0.01	14.57	<0.01%
Cynicism/ Depersonal-ization	17	0.14	0.01	0.27	2.28*	0.04	14.36	<0.01%
Personal Accomplish-ment	19	0.21	0.06	0.35	3.04**	0.06	24.57	26.7%
Total Burnout	6	0.35	-0.02	0.72	2.43	0.07	8.96	44.2%

Note. * statistically significant at $p < .05$; ** statistically significant at $p < .01$; *** statistically significant at $p < .001$, *k* is the number of studies included in the analysis, Hedges *g* is the mean effect size, 95% CI are the minimum and maximum limits of the confidence interval around Hedges *g*, *t* is the statistical test used for computing the statistical significance of the mean effect size when using the HKSJ method, τ^2 is an estimate of the variance of the effect size parameters across the population of studies, *Q* is a statistical test used for the estimation of heterogeneity, I^2 is another statistical test to estimate heterogeneity and represents the percentage of the variance accounted for by heterogeneity rather than error.

Heterogeneity for Emotional Exhaustion Post-Intervention

The individual study effect sizes for Emotional Exhaustion at post-intervention are reported in a forest plot in Figure 2. As seen in table 3, the *Q* statistic was not statistically significant, Q (24) =14.57, p = .9326. This result, combined with examining the I^2 statistic (<0.01%; CI, <0.01%; 7.6%), and the variance of the effect sizes (τ^2 = 0.01) together suggest that the effect sizes for Emotional Exhaustion have small levels of heterogeneity.

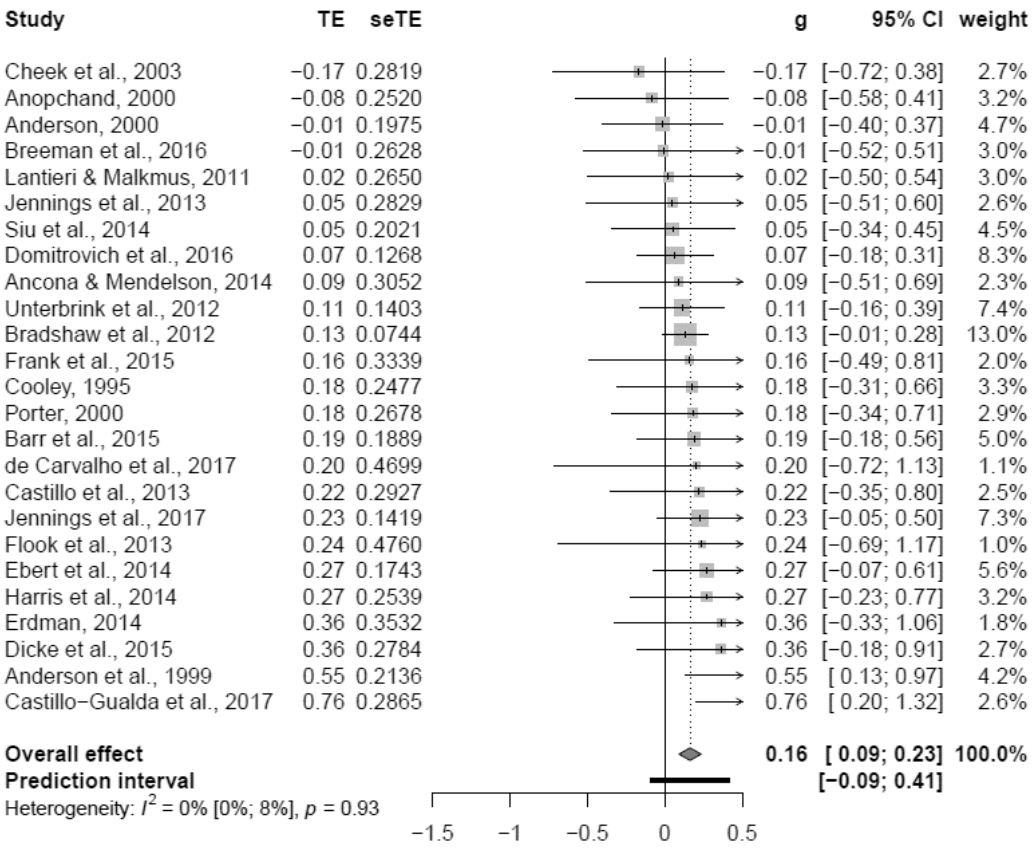

Figure 2 Emotional Exhaustion Post-Intervention Forest Plot

Heterogeneity of Cynicism/Depersonalization

The individual study effect sizes for Cynicism/Depersonalization at post-intervention are reported in a forest plot in Figure 3. Statistical analyses of homogeneity, Q (16) = 14.36, p = .5715 and I^2 <.01%; CI, [<.01%-45.5%] and the between study variance component estimate (τ^2 = 0.04) together suggest a very small level heterogeneity for Cynicism/Depersonalization effect sizes.

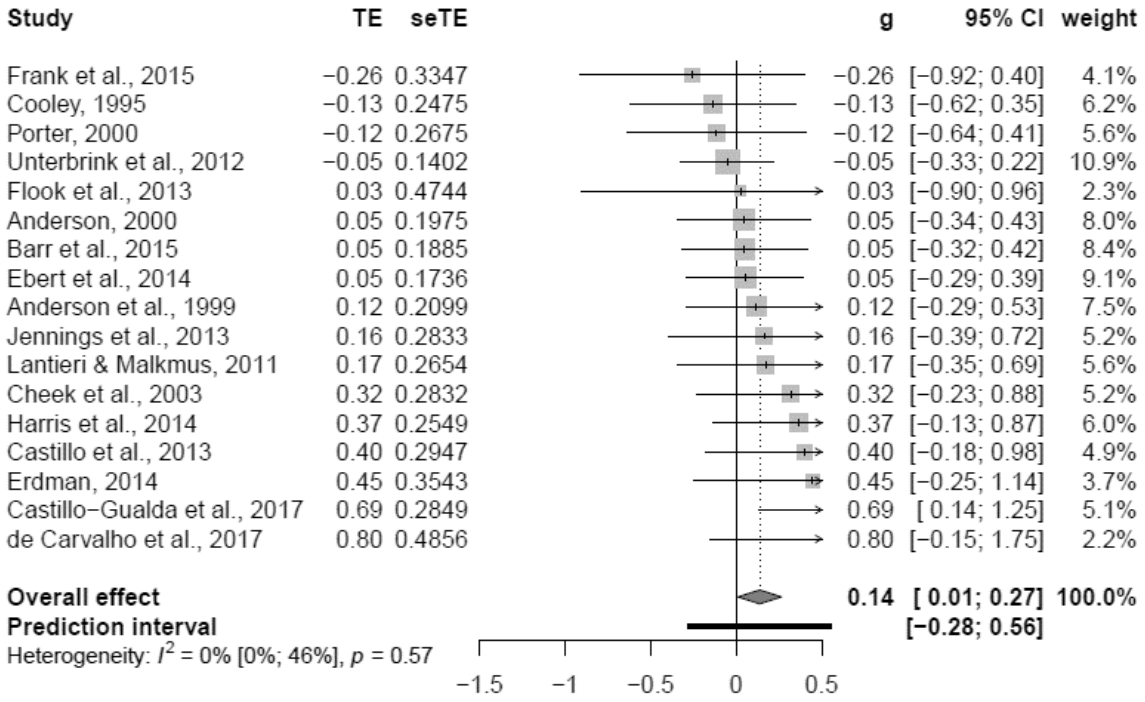

Figure 3 Cynicism/Depersonalization Post-Intervention Forest Plot

Heterogeneity of Personal Accomplishment

The individual study effect sizes for Personal Accomplishment at post-intervention are reported in a forest plot in Figure 4. Examining the our statistical assessments of homogeneity, Q

(18)=24.57, p = .1374), the I^2 statistic (26.7%; CI, [<.01%-58.0%]), and our estimate of he between study variance component(τ^2 = 0.06) suggests a small degree of heterogeneity for Personal Accomplishment effect sizes.

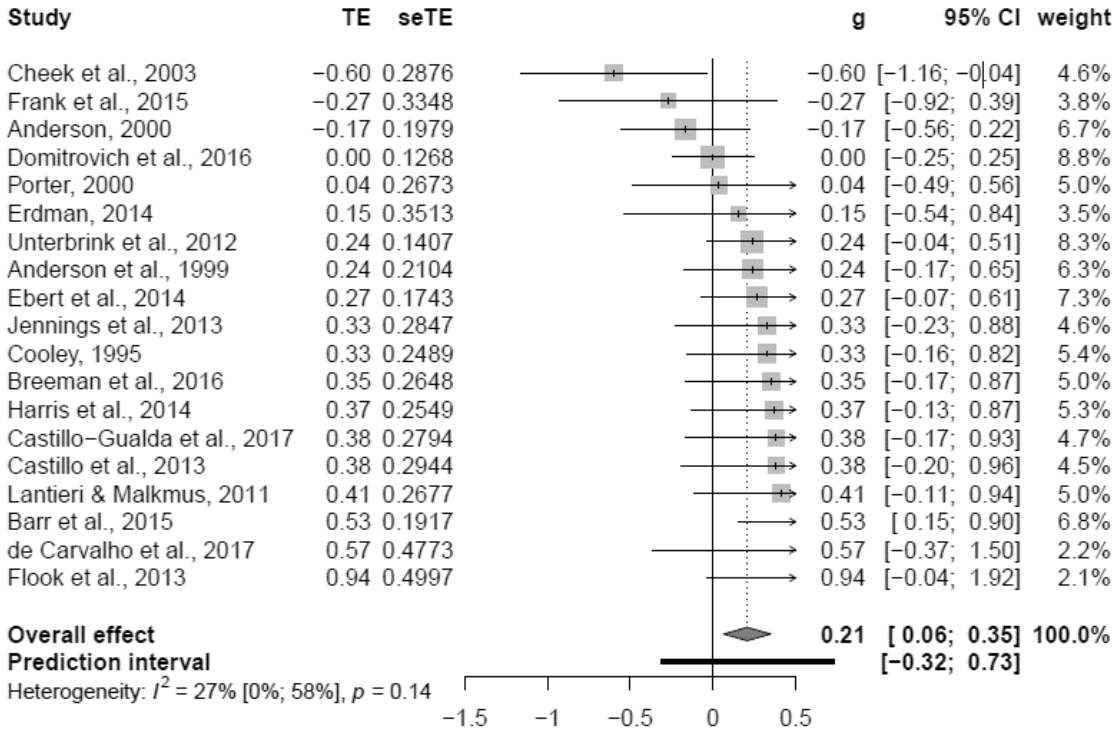

Figure 4 Personal Accomplishment Post-Intervention Forest Plot

Heterogeneity of Total Burnout Symptoms

The individual study effect sizes for Total Burnout Symptoms at post-intervention are reported in a forest plot in figure 5. Examining Q (5) =8.96, p=.1109), the I^2 statistic (44.2%; CI, [<.01%-77.9%]), and the estimate for the between-studies variance component (τ^2 = 0.07) suggests that the effect sizes for Total Burnout symptoms have a small level heterogeneity.

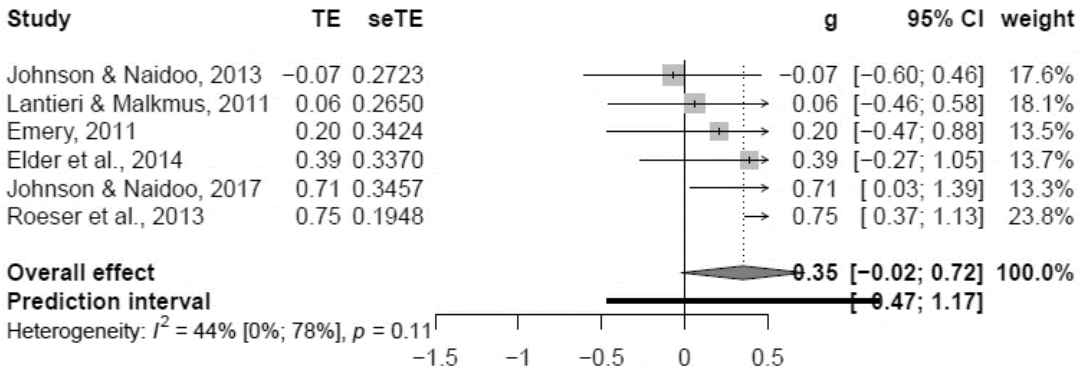

Figure 5 Total Burnout Post Intervention Forest Plot

Heterogeneity of Follow-Up

Considerably fewer included studies collected follow-up compared to post-intervention data (Appendix C). Taken together, the *Q* statistics, I^2 statistics, τ^2, and forest plots for follow-up data suggest low levels of heterogeneity for all three facets of burnout (see table 4 and figure 6, figure 7, and figure 8). Further, only one study included follow-up data for Total Burnout Symptoms, thus, no results are presented for Total Burnout Symptoms at follow-up.

Table 4 Summary Effect Sizes at Follow-up

Scale	*k*	Hedges *g*	95% CI		*t*	τ^2	*Q*	I^2
Emotional Exhaustion	3	0.52	-0.03	1.07	4.06	0.02	1.88	<0.01%
Cynicism/ Depersonalization	2	0.22	-1.01	1.45	2.31	<0.01	0.50	<0.01%
Personal Accomplishment	2	0.29	0.09	0.49	18.13*	<0.01	0.01	<0.01%

Note. * statistically significant at p < .05; ** statistically significant at p < .01; *k* is the number of studies included in the analysis, Hedges *g* is the mean effect size, CI are the minimum and maximum limits of the confidence interval around Hedges *g;* however, is not generated when there are only two studies. *t* is the statistical test used for computing the statistical significance of the mean effect size when using the HKSJ method, τ^2 is an estimate of the variance of the effect size parameters across the population of studies, *Q* is a statistical test used for the estimation of heterogeneity, I^2 is another statistical test to estimate heterogeneity and represents the percentage of the variance accounted for by heterogeneity rather than error.

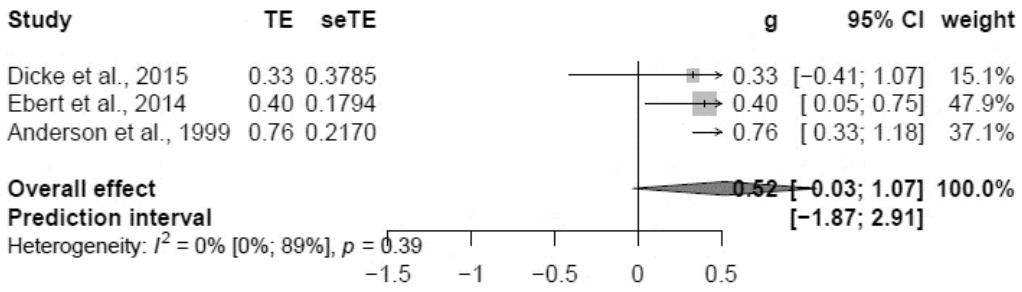

Figure 6 Emotional Exhaustion Follow-Up Forest Plot

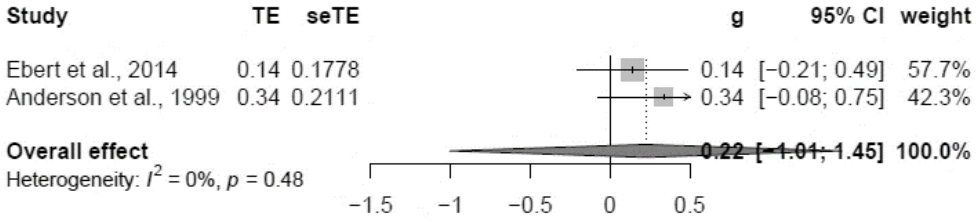

Figure 7 Cynicism/Depersonalization Follow-Up Forest Plot

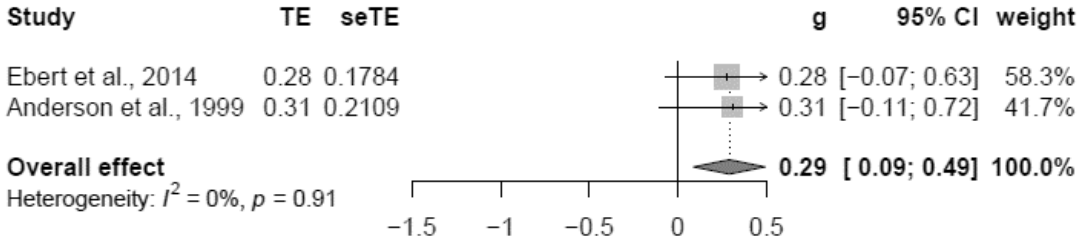

Figure 8 Personal Accomplishment Follow-Up Forest Plot

Analyses of Bias

We examined bias by conducting meta-regressions for each of the facets of burnout with use of an active vs. passive control group (dummy coded) and type of assignment to condition (dummy coded) as moderators. Many studies randomly assigned schools, rather than teachers, to condition either for practical reasons or to avoid contamination of the control groups. Several studies simply used convenience assignment to condition (Castillo et al., 2013; Dicke et al., 2015; Porter, 2000; Siu et al., 2014).

None of the overall models examining bias indicators were statically significant for any of the 3 facets of burnout. However, within the model tested for Personal Accomplishment, the use of an active control group was a statistically significant moderator (β = -0.34, *SE* = 0.13, *p* =.009, 95% CI, [-0.59; -0.08]). Of note, only 4 of the 19 studies that provided Personal Accomplishment data used an active control group. The negative relationship between use of an active control group suggests that studies that used an active control group, rather than a passive or wait-list control group, tended to have smaller effect sizes for Personal Accomplishment. Table 5 presents results for Emotional Exhaustion, table 6 presents results for Cynicism/Depersonalization, and table 7 presents results for Personal Accomplishment.

Table 5 Study Bias Indicators as Moderators for Emotional Exhaustion

Emotional Exhaustion	***k***	***β***	τ^2	I^2	R^2	Q_R	Q_M
Control group type	25	-0.10	<0.01	<0.01%	<0.01%	13.43	1.14
Assignment by school		<0.01					
Assignment by classroom		-0.04					

Note. * = $p < .05$, ** = $p < .01$, *k* = number of studies included in analysis, β = standardized regression coefficient (i.e. the slope of a line in a regression equation), τ^2 = estimated amount of residual heterogeneity, I^2 = percent of estimated true heterogeneity, R^2 = amount of heterogeneity accounted for by the model, Q_R = statistic for residual heterogeneity/goodness of fit, Q_M = chi-squared statistic for testing null hypothesis that none of the covariates are related to effect size, Control group type refers to studies with active vs. passive control groups, Assignment by school refers to studies that used random assignment or a matching design at the school level, Assignment by classroom refers to studies that were randomly assigned or used a matching design and the classroom/teacher level.

Table 6 Study Bias Indicators as Moderators for Cynicism/Depersonalization

Cynicism/ Depersonalization	***k***	***β***	τ^2	I^2	R^2	Q_R	Q_M
Control group type	17	0.39	<0.001	<0.01%	<0.01%	9.10	5.27
Assignment by school		0.30					
Assignment by classroom		0.08					

Note. * = $p < .05$, ** = $p < .01$, *k* = number of studies included in analysis, β = standardized regression coefficient (i.e. the slope of a line in a regression equation), τ^2 = estimated amount of residual heterogeneity, I^2 = percent of estimated true heterogeneity, R^2 = amount of heterogeneity accounted for by the model, Q_R = statistic for residual heterogeneity/goodness of fit, Q_M = chi-squared statistic for testing null hypothesis that none of the covariates are related to effect size, Control group type refers to studies with active vs. passive control groups, Assignment by school refers to studies that used random assignment or a matching design at the school level, Assignment by classroom refers to studies that were randomly assigned or used a matching design and the classroom/teacher level.

Table 7 Study Bias Indicators as Moderators for Personal Accomplishment

Personal Accomplishment	***k***	***β***	τ^2	I^2	R^2	Q_R	Q_M
Active Control Group	19	-0.34**	<0.01	<0.01%	99.98%	17.67	6.90
Assigned by school		0.01					
Assigned by classroom		-0.13					

Note. * = $p < .05$, ** = $p < .01$, k = number of studies included in analysis, β = standardized regression coefficient (i.e. the slope of a line in a regression equation), τ^2 = estimated amount of residual heterogeneity, I^2 = percent of estimated true heterogeneity, R^2 = amount of heterogeneity accounted for by the model, Q_R = statistic for residual heterogeneity/goodness of fit, Q_M = chi-squared statistic for testing null hypothesis that none of the covariates are related to effect size, Control group type refers to studies with active vs. passive control groups, Assignment by school refers to studies that used random assignment or a matching design at the school level, Assignment by classroom refers to studies that were randomly assigned or used a matching design and the classroom/teacher level.

Research Question 1

Research question 1 asked: what is the average weighted effect size for interventions to prevent teacher burnout? We hypothesized that overall, interventions would produce at least a small effect in reducing burnout, as defined using Cohen's conventions (1988). We tested this hypothesis by calculating the summary effect size statistic (g) for all included studies, using a random effects model, using the metagen package in R, with the HKSJ method. The hypothesis for research question 1 was partially supported in that two out of the four summary effect sizes can be considered small by Cohen's conventions (1988).

The mean effect sizes at post-intervention, for each facet of burnout, as well as Total Burnout symptoms, are presented in table 3. The mean effect size for Total Burnout (g = 0.35, 95% CI, [-0.02, 0.72], p = .06) and the mean effect size for Personal Accomplishment (g = 0.21,

95% CI, [0.06; 0.35], $p < .01$) are both considered small by Cohen's conventions (1988). The mean effect sizes for Emotional Exhaustion ($g = 0.16$, 95% CI, [0.09, 0.23], $p < .001$), Cynicism/ Depersonalization ($g = 0.14$, 95% CI, [0.01, 0.27], $p = .037$) is considered trivial by Cohen's conventions (1988). While the effect sizes for all three facets of burnout were statistically significant at alpha = .05, setting alpha at a different level, such as .01, would change the interpretation of statistical significance. For the purposes of this study, our interest lies, not in the statistical significance of the effect sizes, but rather the magnitude of the effect sizes and our interpretation of the practical significance of these effect sizes. Regardless of what alpha is set at, Hedges *g* and the 95% confidence interval around each effect size remains the same and this information is what we weigh most heavily in our interpretation of the mean effect sizes.

As mentioned previously, Cohen's conventions (1988) must be used with caution (Durlak, 2009; Thompson, 2002). Durlak (2009) recommends interpreting effect sizes in the context of the larger literature as well as for the clinical meaningfulness of the effect. These standardized mean difference effect sizes are somewhat larger than those presented by Iancu et al. (2018) in their comparable meta-analysis of teacher burnout interventions, which utilized Cohen's *d* (an unadjusted standardized mean difference score) as the effect size statistic.

Research Question 2

Research question 2 asked: for those studies that included follow-up data, were effects maintained over time? We hypothesized that only trivial differences would remain at follow-up, that is, an average standardized mean difference effect size smaller than .20 (Cohen, 1988). Only one study collected follow-up data for Total Burnout symptoms, and therefore, a meta-analysis on Total Burnout at follow-up was not possible (see table 4). Hypothesis 2 was not supported; contrary to Hypothesis 2, the follow-up effect sizes for Emotional Exhaustion ($g = 0.52$, 95% CI

[-0.03, 1.07], p = .056) were moderate by Cohen's conventions. Also contrary to our hypothesis, follow-up effect sizes for both Cynicism/Depersonalization (g = 0.22, 95% CI,[-1.01, 1.45], p = .478) and Personal Accomplishment (g = 0.29, 95% CI, [-0.09, 0.49], p <.05) are all considered small by Cohen's conventions (1988), rather than trivial.

The sample sizes (k) for all three of these meta-analyses are small, and results should be interpreted with caution. We present the effect sizes for follow-up, in part to encourage future researchers to collect and report this data to increase the universe of studies that measure follow-up data. Further, Cohen's conventions are not the only, nor the best way to interpret magnitude of effect sizes. Table 8 presents the present study follow-up effect sizes side by side to the unadjusted standardized mean effect sizes reported by Iancu et al. (2018) for "1-3 months" and ">3 months" of "time lag" between intervention completion and data collection (p. 386). Of note, Iancu et al. (2018) included many of the same studies in their analyses as the present study; however, some studies did differ due to different search terms and exclusion criteria.

Table 8 Effect Sizes for Follow-up: Present Study vs. Iancu et al. (2018)

Scale	**Present Study *g's***	**Iancu et al., 2018 *d's***	
Emotional Exhaustion	0.52	d = 0.68 d =0.46	"> 3 months" lag "1-3 months" lag
Cynicism/ Depersonalization	0.22	d = 0.16	"1-3 months" lag
Personal Accomplishment	0.29	d = 0.29	"1-3 months" lag
Note. d = Cohen's *d, d's* presented in Iancu et al. (2018, p. 386)			

Research Question 3

Research question 3 asked: Do different types of interventions have stronger effects? Our results are described in the following section. We hypothesized that mindfulness interventions targeting teachers will have larger effects than classroom management or other behavioral interventions. Social Emotional Learning programs, internet-based programs, and psycho-education interventions without a clear mindfulness component were excluded from analyses to test this hypothesis. Out of the 30 total studies included in analyses, 14 studies used a mindfulness intervention, including but not limited to meditation (Ancona & Mendelson, 2014; Anderson et al., 1999; de Carvalho et al., 2017; Elder et al., 2014; Emery, 2011; Flook et al., 2013; Frank et al., 2015; Harris et al., 2014; Jennings et al., 2013; Jennings et al., 2017; Johnson & Naidoo, 2013; Johnson & Naidoo, 2017; Lantieri & Malkmus, 2011; Roeser et al., 2013). Four studies included used a behavioral intervention (PBISplus, GBG, and so on; Bradshaw et al., 2012; Breeman et al., 2016; Dicke et al., 2015; Domitrovich et al., 2016); however, one of these studies (Dicke et al., 2015) also included another intervention group and the classroom management intervention was not randomly selected for inclusion in this meta-analysis. Thus only three studies using a behavioral approach are included in the analyses to test Hypothesis 3. None of the studies that tested a behavioral intervention reported Total Burnout Symptoms or Cynicism/ Depersonalization as an outcome, therefore Hypothesis 3 could only be tested for Emotional Exhaustion and Personal Accomplishment.

Table 9 presents the meta-regression results comparing mindfulness interventions to behavioral interventions. Intervention type did not appear to be a statistically significant moderator for Emotional Exhaustion nor Personal Accomplishment (see table 9). While table 9

presents a high R^2 value (R^2 =99.99%) for Personal Accomplishment, the model was not found to be statistically significant.

Table 9 Meta-regression Results: Mindfulness Interventions vs. Behavioral Interventions

Scale	***k***	***β***	τ^2	I^2	R^2	Q_R	Q_M
Emotional Exhaustion	12	0.11	<0.01	<0.01	<0.01	4.02	1.36
Personal Accomplishment	9	0.24	<0.01	<0.01	99.99%	6.61	2.23
Note. * = $p < .05$, ** = $p < .01$, k = number of studies included in analysis, β = standardized regression coefficient (i.e. the slope of a line in a regression equation), τ^2 = estimated amount of residual heterogeneity, I^2 = percent of estimated true heterogeneity, R^2 = amount of heterogeneity accounted for by the model, Q_R = statistic for residual heterogeneity/goodness of fit, Q_M = chi-squared statistic for testing null hypothesis that none of the covariates are related to effect size,							

Next, we examined the summary effect sizes, confidence intervals, and forest plots for Mindfulness interventions and Behavioral interventions separately, for both Personal Accomplishment and Emotional Exhaustion. For both Emotional Exhaustion (figure 9 and figure 10) and Personal Accomplishment (figure 11 and figure12), the summary effect sizes were higher for Mindfulness interventions, compared with the effect sizes produced by Behavioral interventions. Mindfulness interventions produced statistically significant effect sizes for both Emotional Exhaustion (g = 0.23, 95% CI, [0.10; 0.36], t= 3.98, $p < .01$) and Personal Accomplishment (g = 0.31, 95% CI, [0.0327; 0.5967], t= 2.73, $p < .05$), whereas Behavioral interventions did not (table 10).

Table 10 Mindfulness vs. Behavioral Summary Effect Sizes

	Mindfulness				Behavioral			
Scale	***k***	**Hedges *g***	***95% CI***	***t***	***k***	**Hedges *g***	**95% CI**	***t***
Emotional Exhaustion	9	0.23	[0.0957; 0.3597]	3.98**	3	0.11	[-0.018; 0.2332]	3.68
Personal Accomplishment	7	0.31	[0.0327; 0.5967]	2.73*	2	0.11	[-1.9419; 2.1547]	0.66
Note. * statistically significant at p < .05; ** statistically significant at p < .01; *k* is the number of studies included in the analysis, Hedges *g* is the mean effect size, 95% CI represents the limits of the 95% confidence interval around Hedges *g*, t is the statistical test used for computing the statistical significance of the mean effect size when using the HKSJ method.								

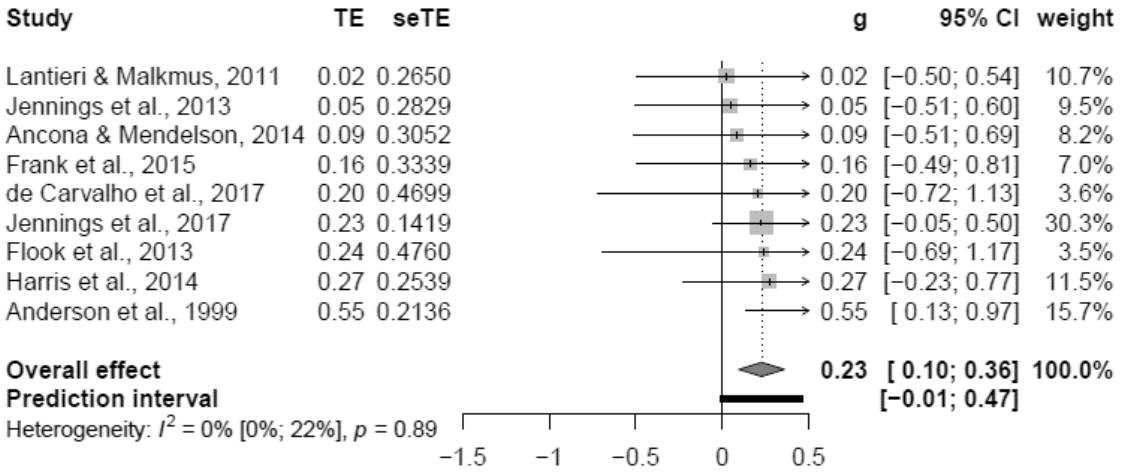

Figure 9 Forest Plot for Emotional Exhaustion Effect Sizes for Mindfulness Interventions

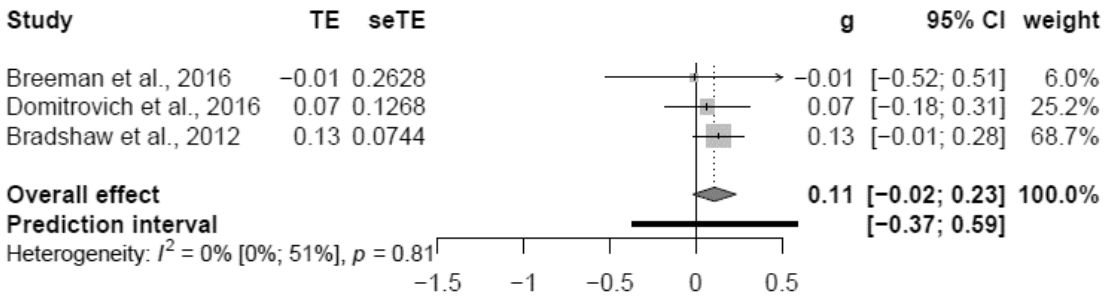

Figure 10 Forest Plot for Emotional Exhaustion for Behavioral Interventions

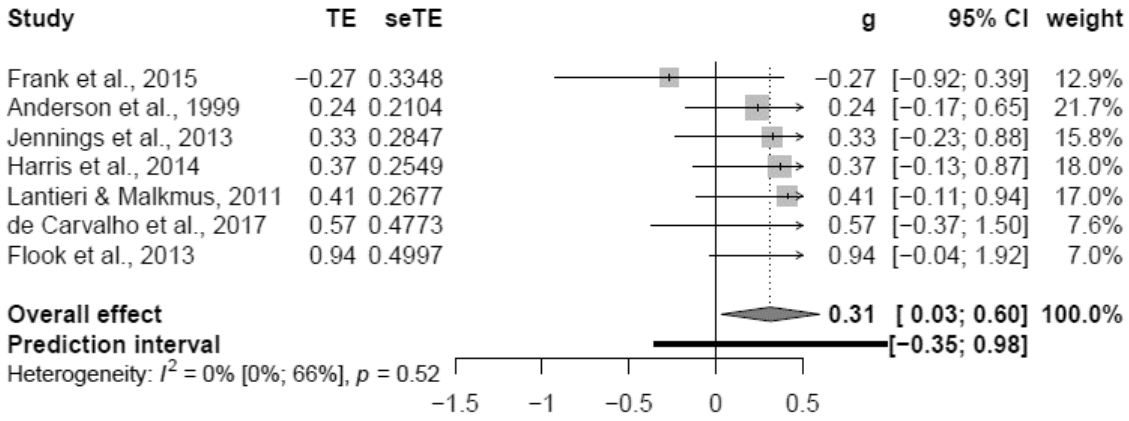

Figure 11 Forest Plot for Personal Accomplishment for Mindfulness Interventions

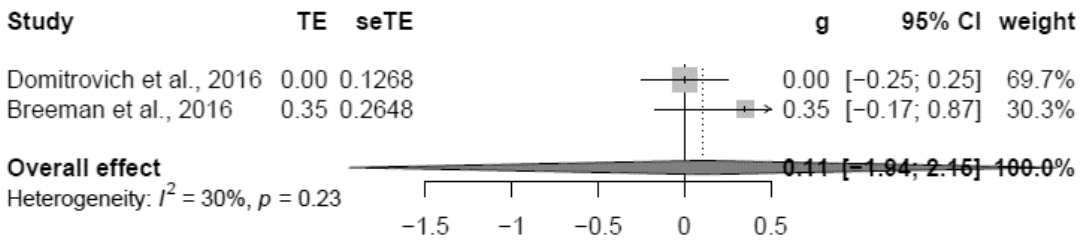

Figure 12 Forest Plot for Personal Accomplishment Effect Sizes for Behavioral Interventions

Research Question 4

Research question 4 asked: what teacher and school characteristics moderate intervention effectiveness? We hypothesized that many factors, including length of intervention, use of homework, work experience, age, sex, race, and student socio-economic status (SES), may moderate effectiveness. We tested eight specific hypotheses described below.

Hypothesis 4a

We hypothesized that studies with interventions implemented over longer periods of time will produce larger effect sizes for each facet of burnout. To test this hypothesis, we measured weeks of direct contact with interventionists (including latency between contacts with interventionists) and total minutes of intervention (minutes spent with an interventionist in psychoeducation, training, and so on) and examined both weeks of direct contact and minutes of direct contact as independent moderators for all facets of burnout.

Of the 5 studies that included Total Burnout Symptoms as an outcome, only three reported the weeks of direct contact or minutes of direct contact; thus, hypothesis 4a was not tested for Total Burnout Symptoms. Weeks of direct contact did not appear to moderate

intervention effectiveness for any of the facets of burnout (see table 11). Minutes of direct contact was a statistically significant moderator of intervention effects on Personal Accomplishment (R^2= 83.25%, β <.01, Q_M (1)= 7.04, p<0.01). While the slope of the regression line appears small, this is not unexpected given that the continuous moderator is scaled in minutes. Minutes of direct contact was not a statistically significant moderator for the other facets of burnout (see table 12).

Table 11 Weeks of Direct Contact as a Moderator

Scale	***k***	***β***	τ^2	I^2	R^2	Q_R	Q_M
Emotional Exhaustion	22	>-0.01	<0.01	<0.01%	<0.01%	13.92	0.37
Cynicism/ Depersonalization	14	>-0.01	<0.01	<0.01%	<0.01%	11.80	0.36
Personal Accomplishment	16	<0.01	0.0150	21.46%	<0.01%	19.04	0.55
Note. * = p < .05, ** = p < .01, k = number of studies included in analysis, β = standardized regression coefficient (i.e. the slope of a line in a regression equation), τ^2 = estimated amount of residual heterogeneity, I^2 = percent of estimated true heterogeneity, R^2 = amount of heterogeneity accounted for by the model, Q_R = statistic for residual heterogeneity/goodness of fit, Q_M = chi-squared statistic for testing null hypothesis that none of the covariates are related to effect size							

Table 12 Minutes of Direct Contact as a Moderator

Scale	k	β	τ^2	I^2	R^2	Q_R	Q_M
Emotional Exhaustion	25	>-0.01	<0.01	0.01%	<0.01%	17.68	0.01
Cynicism/ Depersonalization	15	<0.01	<0.01	<0.01%	51.00%	12.71	0.49
Personal Accomplishment	17	<0.01	<0.01	5.89%	83.25%	15.82	7.04**

Note. * = $p < .05$, ** = $p < .01$, k = number of studies included in analysis, β = standardized regression coefficient (i.e. the slope of a line in a regression equation), τ^2 = estimated amount of residual heterogeneity, I^2 = percent of estimated true heterogeneity, R^2 = amount of heterogeneity accounted for by the model, Q_R = statistic for residual heterogeneity/goodness of fit, Q_M = chi-squared statistic for testing null hypothesis that none of the covariates are related to effect size

Hypothesis 4b

We hypothesized interventions that incorporating homework would produce larger effects for each facet of burnout. To test this hypothesis, we measured weeks of homework assigned to participants and performed a random effect meta-regression with weeks of homework as a continuous moderator. Meta-regression is not recommended when $k< 10$, thus, Emotional Exhaustion was the only outcome used to test Hypothesis 4b. Hypothesis 4b was not supported as homework was not a statistically significant moderator for Emotional Exhaustion effect sizes ($\beta = 0.01$, $R^2 < 0.01\%$, $Q_M(1) = 0.43$, $p = 0.5118$). Table 13 presents results of the meta-regression for homework as a potential moderator of Emotional Exhaustion; weeks of homework did not significantly moderate intervention effectiveness for Emotional Exhaustion.

Homework was operationally defined as clear, explicit instructions or encouragement to engage in a cognitive, affective, or behavioral task (e.g. meditation, carrying out a curriculum, and so on) outside of the scheduled intervention hours or direct contact hours with the

interventionist. Online writing tasks that occurred within online programs were not considered homework and were instead coded as intervention time (not direct contact with an interventionist). Some intervention studies that used the Good Behavior Game, Social Emotional Interventions, or other curricula, appeared to contain homework based on our definition of homework, in that teachers needed to prepare and carry out an intervention in their classes; however, if this homework was not specified in the methods of these studies, weeks of assigned homework was coded as "9999" (e.g., not specified) and these studies were excluded from analyses testing this hypothesis. Other interventions, such as Expressive Writing (Anchopand, 2000), clearly did not contain a homework component, therefore "weeks of homework" was coded as "0" and these studies were included in analyses.

Table 13 Weeks of Homework as a Moderator

Scale	***k***	***β***	τ^2	I^2	R^2	Q_R	Q_M
Emotional Exhaustion	11	0.01	<0.01	<0.01%	<0.01%	5.84	0.43
Note. * = $p < .05$, ** = $p < .01$, *k* = number of studies included in analysis, β = standardized regression coefficient (i.e. the slope of a line in a regression equation), τ^2 = estimated amount of residual heterogeneity, I^2 = percent of estimated true heterogeneity, R^2 = amount of heterogeneity accounted for by the model, Q_R = statistic for residual heterogeneity/goodness of fit, Q_M = chi-squared statistic for testing null hypothesis that none of the covariates are related to effect size							

Hypothesis 4c

We hypothesized that samples with a lower mean for years of work experience would see a greater reduction in burnout symptoms for all three facets of burnout. Table 14 presents our results from the meta-regressions on Emotional Exhaustion, Cynicism/Depersonalization, and

Personal Accomplishment. Mean years of work experience was not a statistically significant moderator for any facet of burnout. Total Burnout symptoms was excluded from these analyses due to low sample size.

Table 14 Mean Years Work Experience as a Moderator

Scale	***k***	***β***	τ^2	I^2	R^2	Q_R	Q_M
Emotional Exhaustion	15	<0.01	<0.01	<0.01%	<0.01%	8.39	0.02
Cynicism/ Depersonalization	11	-0.01	<0.01	<0.01%	<0.01%	9.22	0.64
Personal Accomplishment	12	0.01	0.06	46.08%	<0.01%	18.56*	0.05

Note. * = $p < .05$, ** = $p < .01$, k = number of studies included in analysis, β = standardized regression coefficient (i.e. the slope of a line in a regression equation), τ^2 = estimated amount of residual heterogeneity, I^2 = percent of estimated true heterogeneity, R^2 = amount of heterogeneity accounted for by the model, Q_R = statistic for residual heterogeneity/goodness of fit, Q_M = chi-squared statistic for testing null hypothesis that none of the covariates are related to effect size

Hypothesis 4d

We hypothesized that samples with a lower mean age would see greater reductions in Emotional Exhaustion and Cynicism/Depersonalization compared to samples with lower proportions of younger teachers. Hypothesis 4d was not supported. Mean age was not a statistically significant moderator for Emotional Exhaustion nor Cynicism/Depersonalization (see table 15).

Table 15 Mean Age as a Moderator

Scale	k	β	τ^2	I^2	R^2	Q_R	Q_M
Emotional Exhaustion	14	>-0.001	<0.01	<0.01	>0.01%	7.03	0.07
Cynicism/ Depersonalization	10	-0.02	<0.01	9.30%	<0.01%	9.42	0.85
Note. * = $p < .05$, ** = $p < .01$, k = number of studies included in analysis, β = standardized regression coefficient (i.e. the slope of a line in a regression equation), τ^2 = estimated amount of residual heterogeneity, I^2 = percent of estimated true heterogeneity, R^2 = amount of heterogeneity accounted for by the model, Q_R = statistic for residual heterogeneity/goodness of fit, Q_M = chi-squared statistic for testing null hypothesis that none of the covariates are related to effect size							

Hypothesis 4e

We hypothesized that samples with higher proportions of female teachers would see a greater reduction in Emotional Exhaustion compared to samples with lower proportions of female teachers. Contrary to hypothesis 4e, the percentage of female teachers in each study was not associated with effect sizes for Emotional Exhaustion (table 16).

Table 16 Percent Female as a Moderator

Scale	k	β	τ^2	I^2	R^2	Q_R	Q_M
Emotional Exhaustion	20	>-0.01	<0.01	<0.01%	<0.01%	13.80	0.03
Note. * = $p < .05$, ** = $p < .01$, k = number of studies included in analysis, β = standardized regression coefficient (i.e. the slope of a line in a regression equation), τ^2 = estimated amount of residual heterogeneity, I^2 = percent of estimated true heterogeneity, R^2 = amount of heterogeneity accounted for by the model, Q_R = statistic for residual heterogeneity/goodness of fit, Q_M = chi-squared statistic for testing null hypothesis that none of the covariates are related to effect size,							

Hypothesis 4f

We hypothesized that samples with higher proportions of male teachers would see a greater reduction in Cynicism/Depersonalization compared to samples with lower proportions of male teachers; however, only 9 studies that reported Cynicism/Depersonalization results also reported the percentage of male teachers. Meta-regression is not recommended for fewer than 10 studies, thus, we did not test this hypothesis.

Hypothesis 4g

We hypothesized that samples with higher proportions of White teachers will show greater reduction in burnout symptoms across domains compared to samples with lower proportions of White teachers. Insufficient data were available to test Hypothesis 4g for Cynicism/Depersonalization, Personal Accomplishment, and Total Burnout symptoms. For Emotional Exhaustion, hypothesis 4g was not supported as the percent of White participants was not a statistically significant moderator (table 17).

Table 17 Percent White as a Moderator

Scale	***k***	***β***	τ^2	I^2	R^2	Q_R	Q_M
Emotional Exhaustion	10	>-0.01	<0.01	<0.01%	<0.01%	5.50	<0.01

Note. * = $p < .05$, ** = $p < .01$, k = number of studies included in analysis, $\beta_=$ standardized regression coefficient (i.e. the slope of a line in a regression equation), τ^2 = estimated amount of residual heterogeneity, I^2 = percent of estimated true heterogeneity, R^2 = amount of heterogeneity accounted for by the model, Q_R = statistic for residual heterogeneity/goodness of fit, Q_M = chi-squared statistic for testing null hypothesis that none of the covariates are related to effect size

Hypothesis 4h

We hypothesized that samples conducted in schools with lower proportions of students with free and reduced lunch, that is, schools with a majority high SES population, will show greater reductions in burnout symptoms across domains. Few data on free and reduced lunch were available, thus, studies that reported students were "majority low SES," "majority high SES," or "majority middle SES" were coded categorically.

Nine studies included information about the overall socio-economic status of the students in the schools. These data were coded categorically, such that studies conducted on teachers in majority high or middle SES schools can be compared to studies whose participants taught in majority low SES schools. The only studies that collected Total Burnout symptoms data were both conducted with low SES student populations (Johnson & Naidoo, 2013; 2017), thus, this hypothesis was not tested on Total Burnout symptoms. We did not conduct meta-regressions or an analog to ANOVA for SES due to low study sample sizes; however, statistics for subgroups of SES are presented in Table 18. Figures 13 and 14 are forest plots for Emotional Exhaustion effect sizes for Low SES and High/Middle SES, respectively. Figures 15 and 16 are forest plots for Cynicism/Depersonalization effect sizes for Low SES and High/Middle SES, respectively, and Figures 17 and 18 are forest plots for Personal Accomplishment effect sizes for Low SES and High/Middle SES, respectively. It is not clear whether SES may moderate intervention effectiveness.

Table 18 Socio-Economic Status (SES) as a Moderator

	Low SES					High/Middle SES				
Scale	***k***	**Hedges *g***	***95% CI***	***t***	***p***	***k***	**Hedges *g***	**95%CI**	***t***	***p***
Emotional Exhaustion	4	0.20	[0.12; 0.28]	7.86**	*p* = .004	3	0.45	[-0.31; 1.20]	2.54	*p* =.126
Cynicism/ Depersonalization	2	0.05	[-0.06; 0.15]	5.71	*p* = .110	3	0.56	[0.01; 1.11]	4.35*	*p* = .049
Personal Accomplishment	2	0.60	[-1.40; 2.60]	3.80	*p* = .164	3	0.40	[0.20; 0.60]	8.4*	*p* = .014
Note. * statistically significant at p < .05; ** statistically significant at p < .01; *k* is the number of studies included in the analysis, Hedges *g* is the mean effect size, 95% CI represents the limits of the 95% confidence interval around Hedges *g*, t is the statistical test used for computing the statistical significance of the mean effect size when using the HKSJ method.										

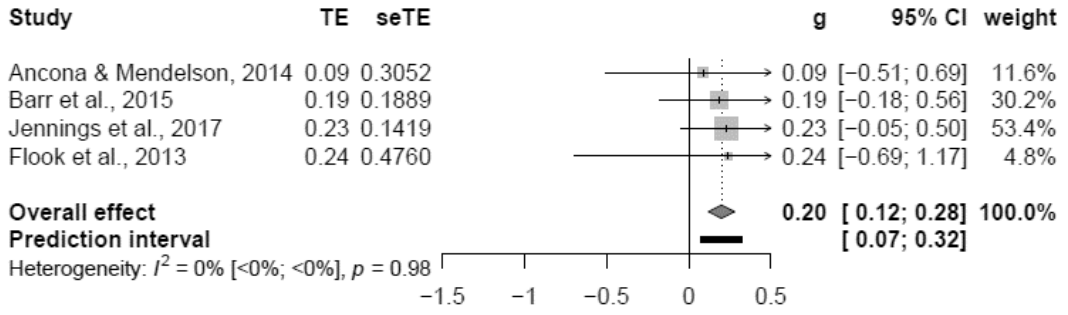

Figure 13 Forest Plot Emotional Exhaustion Low SES

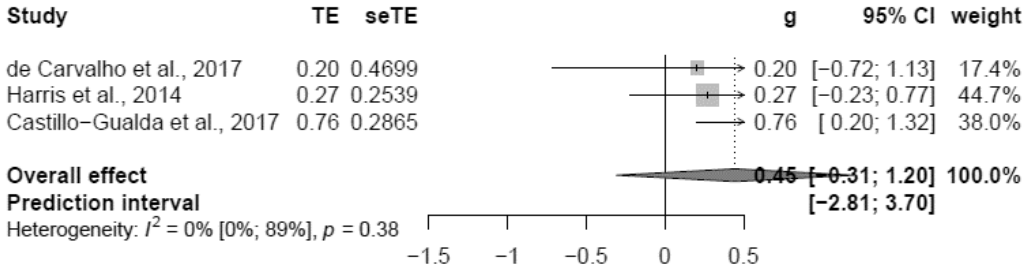

Figure 14 Forest Plot for Emotional Exhaustion High/Middle SES

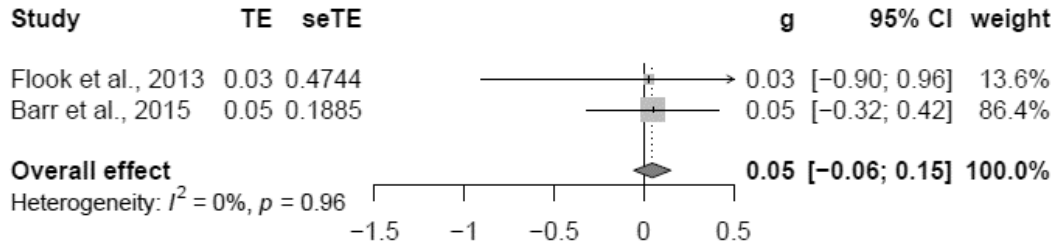

Figure 15 Forest Plot for Cynicism/Depersonalization Low SES

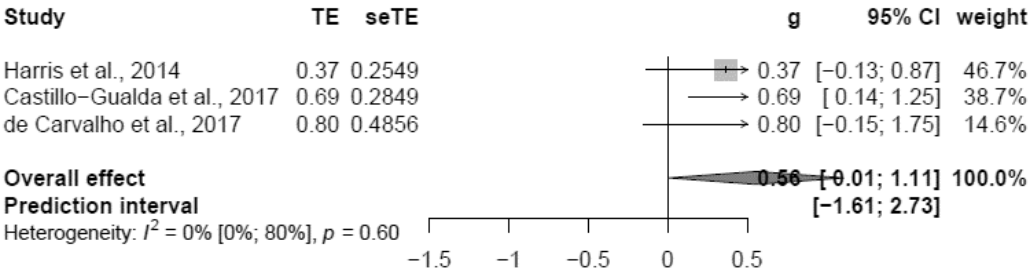

Figure 16 Forest Plot for Cynicism/Depersonalization High/Middle SES

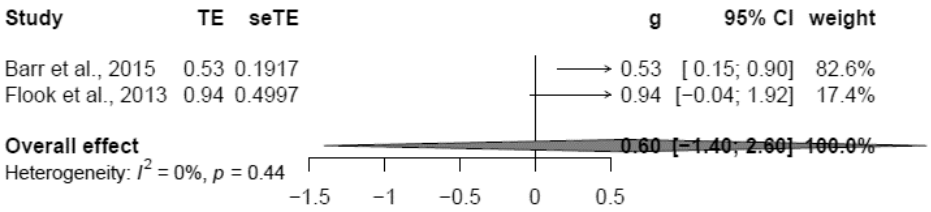

Figure 17 Forest Plot for Personal Accomplishment Low SES

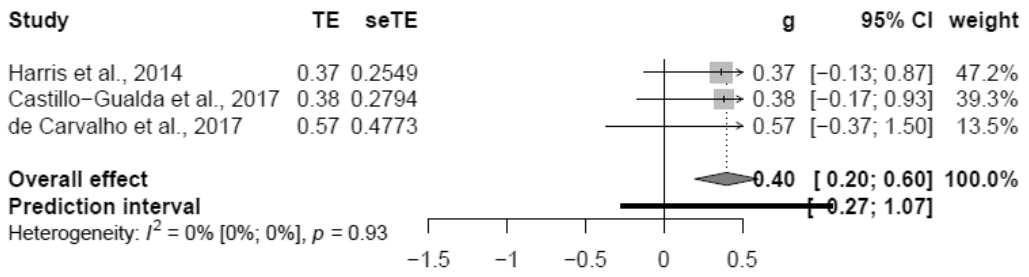

Figure 18 Forest Plot for Personal Accomplishment High/Middle SES

Study Quality

Study quality was measured by the study quality indicator tool in Appendix A; this measure was developed for this study and was based on standards developed by Gersten et al. (2005). Interrater reliability for the study quality indicator alone was initially 79.37%. After deleting and revising items with low inter-rater reliability, reliability was 89.63%. The first author determined the final codes for each item before conducting analyses.

We performed meta-regressions for Emotional Exhaustion, Cynicism/Depersonalization, Personal Accomplishment, and Total Burnout symptoms, with study quality scores as a continuous moderator, despite a small number of studies reporting Total Burnout symptoms. Results indicated study quality was positively associated with Total Burnout effect sizes (β = 0.07, SE = 0.03, R^2> *99%*, Q_M (1)= 7.30, p < .01); however, this result is based on only 6 studies and conducting meta-regressions when k<10 is not recommended; thus these results should be interpreted very cautiously (Borenstein et al., 2009). Study quality was not a statistically significant moderator for other facets of burnout (Table 19).

Table 19 Study Quality as a Moderator

Scale	*k*	β	τ^2	I^2	R^2	Q_R	Q_M
Emotional Exhaustion	25	<0.01	<0.01	0.01%	<0.01%	14.38	0.19
Cynicism/ Depersonalization	17	-0.01	<0.01	<0.01%	<0.01%	14.32	0.05
Personal Accomplishment	19	0.02	0.02	23.55%	0.51%	23.51	0.65
Total Burnout	6	0.07	<0.01	<0.01	>99%	1.65	7.30**

Note. * = $p < .05$, ** = $p < .01$, *k* = number of studies included in analysis, β = standardized regression coefficient (i.e. the slope of a line in a regression equation), τ^2 = estimated amount of residual heterogeneity, I^2 = percent of estimated true heterogeneity, R^2 = amount of heterogeneity accounted for by the model, Q_R = statistic for residual heterogeneity/goodness of fit, Q_M = chi-squared statistic for testing null hypothesis that none of the covariates are related to effect size,

CHAPTER V

DISCUSSION AND CONCLUSIONS

The purpose of the present study was to perform a systematic meta-analysis on interventions to prevent and treat teacher burnout. The present study has expanded search criteria compared to Iancu et al. (2018), measured study quality in a different way (Appendix A) and examined additional potential moderators of intervention effectiveness. These nine potential moderators included factors associated with burnout or associated with related constructs such as turnover or teacher shortages, including: use of time spent in direct contact with interventionists, use of homework, mean years' work experience, mean age, binary sex, percentage of White participants, and study quality summed scores. We also compared the effect sizes for mindfulness interventions vs. behavioral interventions.

Overall, interventions to prevent and reduce teacher burnout do appear to have some evidence of efficacy. Summary effect sizes for Personal Accomplishment, and Total Burnout were small by Cohen's standards (1988), and the summary effect sizes for Emotional Exhaustion and Cynicism/Depersonalization were trivial by Cohen's conventions (1988); however, Cohen's heuristics are not the best method by which to judge an effect size and it is difficult to interpret the practical significance of these effect sizes (Durlak, 2009). Summary effect sizes for follow-up data were larger than expected; all three facets of burnout produced small to medium effect sizes at follow-up compared to the trivial effect sizes expected, again, by Cohen's conventions. Summary effect sizes at follow-up appeared similar to those presented by Iancu et al. (2018). The number of studies that provided follow-up data was small and should be interpreted

cautiously; these results could be interpreted to mean that intervention effects are not only maintained over time but improve over time. Alternatively, it could be that studies that collected follow-up data tended to have larger effect sizes, as this was not examined specifically for post-intervention results.

Results of the meta-regressions comparing mindfulness and behavioral interventions did not yield statistically significant results; however, evaluating the mindfulness and behavioral interventions separately demonstrated that mindfulness interventions had somewhat larger summary effect sizes and somewhat smaller degrees of between-study heterogeneity compared to behavioral interventions. The number of studies included in these analyses were small, and results were limited to Emotional Exhaustion and Personal Accomplishment effect sizes due to missing data. These differences in effect sizes between behavioral interventions and mindfulness interventions are similar to what Iancu et al. (2018), found in that, for both the present study and Iancu et al. (2018), mindfulness interventions produced statistically significant summary effect sizes for Emotional Exhaustion and Personal Accomplishment, whereas the "professional development" (p. 386) interventions (i.e. behavioral interventions) did not. Zarate et al. (2019) also found mindfulness interventions had a statistically significant, small effect on burnout.

Country to our hypothesis and Iancu et al.'s (2018) findings, we did not find length of intervention (i.e. span of time over which direct contact occurred) to moderate intervention effectiveness. We did find; however, that minutes of direct contact appeared as a statistically significant moderator for Personal Accomplishment. If we assume that Personal Accomplishment is an important facet of burnout as the traditional conceptualizations suggest (Maslach & Jackson, 1981; Maslach et al., 2001), then time spent in direct contact with an interventionist may be an important component of effective burnout interventions.

Weeks of homework, mean years' work experience, mean age, percentage of female participants, percentage of White participants did not appear as statistically significant moderators, contrary to hypotheses. While previous research has found that the demographic factors including work experience, age, race, and SES are related to burnout symptoms or related constructs such as turnover, the present study suggests that they do not moderate intervention effectiveness (Achinstein et al., 2010;Brewer & Shapard, 2004; Edmonson & Thompson, 2000; Gavish & Friedman, 2010; Ingersol & May, 2016; Loeb et al., 2005; Sutcher et al., 2016). Weeks of homework did also not appear to moderate intervention effectiveness for Emotional Exhaustion.

It is important to note that while our aims were similar to those of Iancu et al., (2018) and Zarate et al. (2019), our methods were not identical to either. The latter was a meta-analysis of mindfulness only interventions for teachers, which examined burnout along with other constructs including mindfulness, stress, anxiety, and depression (Zarate et al., 2019). Iancu et al., (2018) categorized interventions differently and included school level (e.g. elementary, middle, and high school) and time lag (i.e. latency between intervention and post-test) as moderators. Further, our expanded search terms and criteria yielded an initial pool of 6,287 records through databases and 60 potential articles through other sources, which is over six times the number of records yielded by Iancu et al.'s (2018) criteria and search that yielded an initial pool of 1020 records. Our final sample included 30 distinct records whereas Iancu et al. (2018) included a total of 23 total records. Additionally, our exclusion criteria led us to exclude "Wolf (2015)" due to issues with data collection that caused the post-intervention group to contain participants not included at pre-intervention, whereas Iancu et al., (2018) included this study. Similarly, we excluded "Berg, Bradshaw, Jo, & Ialongo, (2016)" for presenting results from the same sample used in

"Domitrovitch et al. (2016)"; Iancu et al. (2018) included both studies and it is not clear if Iancu et al. (2018) accounted for the overlap in samples in their analyses. We also excluded Cheon et al. (2014) as their measure of burnout, the Athlete Burnout Questionnaire (ABQ; Raedeke & Smith, 2001) appeared to measure a different construct (physical and emotional exhaustion), whereas Iancu et al. (2018) included this study in their analysis.

Implications

Much of the experimental and quasi-experimental research on teacher burnout uses mindfulness interventions and these interventions have evidence of producing statistically significant effect sizes which are small or moderate by Cohen's conventions. The practical significance of the effect sizes for interventions overall is difficult to interpret and thus, including social validity data should be an important consideration for future research. Overall, it is our belief that "small" effect sizes should not deter school systems from incorporating potentially efficacious interventions to support teachers' mental health and reduce symptoms of burnout. When it comes to improving mental health in service professions, "small" effects by Cohen's conventions may be quite meaningful for quality of life, or for related constructs such as teacher/student relationships. Further, we found some preliminary evidence that effects are not only maintained over time, but improve over time, as effect sizes were larger for follow-up data; however, sample sizes for follow-up data were small and all conclusions made regarding follow-up data are tentative and should be interpreted cautiously.

Many types of interventions have some evidence of effectiveness, with mindfulness interventions having some of the strongest empirical support. While a variety of promising interventions for burnout deserve additional research, practitioners and school systems should consider the interventions with the strongest support for their mission and goals. In the case of

choosing interventions for the purposes of reducing burnout, mindfulness-based interventions appear to be one choice with a relatively large number of studies to provide evidence of effectiveness.

Time spent in direct contact with an interventionist may be an important factor in improving Personal Accomplishment. This finding suggests that resources may be best spent on interventions that include a direct contact component with a trained professional. Very short trainings, or one-time workshops, appear less likely to be effective for the purposes of reducing burnout (Iancu et al., 2018).

While examining study quality, we found that our study quality indicator was a statistically significant moderator for Total burnout; however, this result is based off a meta-regression of only 6 studies which is below Borenstein et al.'s (2009) recommended minimum of 10 studies for meta-regression; and thus, we must interpret this result with caution. Study quality may be important for reducing burnout symptoms overall; however, more research is necessary in order to determine if this is the case and what aspects of study quality may be important (e.g. random assignment to condition, progress monitoring, inclusion of social validity, including psychotherapists as interventionists, administering interventions with fidelity, and so on).

Limitations and Future Directions

Perhaps the most important limitations concern the statistical analyses used in the present study. Firstly, the present study uses univariate modeling when multivariate modeling would be more appropriate given the overlap in samples between each scale and shared variance between the facets of burnout (Becker, 2000; Jackson et al., 2010). Our analyses did not utilize correlations between related variables in order to construct a multivariate model and thus, the

overlap between variables is not accounted for in each individual meta-regression. Future research should consider employing multi-variate meta-analysis by including data from the correlation matrices on the coding sheet and using these correlations (or imputed correlations from other burnout literature if correlation matrices are not presented) in order to create a multivariate model.

Another serious limitation is a large degree of missing data from included studies. We used listwise deletion of a large portion of studies for each analysis, which both reduced the sample sizes (k) and presents a problem for obtaining accurate, unbiased results. Newer methods for imputing missing data for meta-analyses specifically are being developed and present a valuable area for future research.

It is important to note that, because of the previously mentioned limitations, the present study included 23 separate meta-regressions, 3 of which contained 3 covariates in the model, the remainder of which only contained one covariate per model. We maintained an alpha level of .05 throughout analyses. Our inclusion of multiple analyses with this alpha level increases the risk of Type 1 error, that is, finding statistically significant results by chance rather than resulting from meaningful differences at the population level. Alternatively, it reduces Type 2 error or the potential to miss meaningful differences.

In addition to missing data, sample sizes for follow-up data were small and no firm conclusions can be made regarding whether effects improve over time. We encourage researchers to include follow-up data in their intervention research. It is our hope that future meta-analyses have a larger pool of studies that examined follow-up data and can draw more firm conclusions.

The “file drawer” problem describes a common issue in meta-analyses where studies with non-statistically significant results do not become published or publicly available. We attempted to mitigate the file drawer problem by maintaining broad search terms and including unpublished works such as theses, dissertations, and funded reports along with published works; however, it is likely that some studies were inadvertently left out. Further, we excluded Wolf et al. (2015) and Hopman et al. (2018) due to issues with their data collection, and thus, we know at least two relevant studies were excluded from our sample.

Finally, the outcome measures we included in our analyses are not without critique. The MBI and MBI-ES do have evidence of yielding reliable data and evidence of an adequate factor structure (Maslach & Jackson, 1981; Maslach et al., 1996); however, these inventories are often criticized for the way items are worded. Each dimension uses only positively or negatively worded items and wording items in this way presents a large problem from a psychometric standpoint as it can lead to artificial factor solutions (Demerouti et al., 2001).

Further, the idea that burnout is best conceptualized as a tripartite model independent from burnout has also come under fire. Schonfeld et al. (2019a) used an exploratory structural equation modeling bifactor analytic approach to examine burnout, depression, and anxiety scales and found that the MBI Emotional Exhaustion scale loaded highly onto a nonspecific psychological distress factor. This research calls into question whether the construct of burnout is truly a tripartite construct distinct from anxiety and depression (Schonfeld et al., 2019a, 2019b). Schonfeld et al. (2019a, 2019b) suggest that the high correlations between measures of burnout and depression indicate that burnout is depressive condition. Schonfeld et al. (2019b) recommends that occupational health researchers shift focus away from burnout and towards depression, in part because depression can be diagnosed by clinicians, and because the term

depression has a more serious connotation that may promote help-seeking behavior when identified and labeled appropriately; however, one could also argue that non-diagnostic labels such as burnout have a place in the mental health lexicon, particularly for describing subclinical issues and/or reducing stigma related to mental health topics. Regardless, we recommend future research on teacher mental health include other outcomes in place of, or in addition to burnout, including depression and anxiety. Future intervention research and meta-analyses should also consider examining these teacher outcomes, as well as attrition, and self-efficacy.

We encourage researchers conducting future meta-analyses to examine additional moderators. While we attempted to examine sex, age, race, experience, type of intervention, direct contact, and homework as potential moderators, missing data precluded some analyses. In future research, it may be important to measure these variables in different ways (e.g. diverse gender identity instead of binary sex, homework completed vs. assigned, direct contact prescribed vs. attended, intersectionality between races, ethnicities, and genders). Further, additional variables should be examined as potential moderators; particularly those found to associate with burnout or related constructs. Iancu et al. (2018) examined teaching level and found that interventions with primary and middle school teachers reported smaller than average effect sizes compared to studies that included mixed-level samples. Geographic region, including the country or classification of the type of setting/population density (i.e. urban, suburban, and rural) may be important variables to consider in future research.

It was beyond the scope of the present study to examine interventions that prevent teacher turnover or attrition, yet this research is also needed. The most recent systematic review of the experimental literature on mentoring interventions was conducted over 10 years ago (Ingersoll & Kralik, 2004). Ingersoll and Kralik (2004) concluded there is likely some benefit of mentoring

interventions on teacher retention and well-being; however, only 10 experimental or quasi-experimental studies were reviewed, and authors cautioned that effects may be minimal and may be drastically impacted by the nature of the mentoring intervention and type of mentor. Thus, future research should examine the effect of mentoring and other interventions on teacher attrition as well as teacher mental health outcomes.

Conclusion

In conclusion, the present study suggests that interventions to prevent and treat teacher burnout have small effects on Total Burnout symptoms and Personal Accomplishment and statistically significant, yet trivial, effect sizes for Emotional Exhaustion and Cynicism/ Depersonalization. It is difficult to interpret the practical significance of these effects. Heterogeneity for all facets of burnout appeared low at post-intervention and follow-up. Summary effect sizes at follow-up were moderate for Emotional Exhaustion, and small for Cynicism/Depersonalization and Personal Accomplishment. Sample sizes were small for follow-up data. The results of various independent, univariate meta-regressions suggest that minutes of direct contact moderated effect sizes for Personal Accomplishment. While the results of the meta-regression comparing mindfulness and behavioral interventions was not statistically significant, mindfulness interventions do appear to have more, and stronger evidence compared to behavioral interventions for the purposes of reducing burnout symptoms.

REFERENCES

References marked with an * were included in the meta-analyses

Achinstein, B., Ogawa, R. T., Sexton, D., & Freitas, C. (2010). Retaining teachers of color: A pressing problem and a potential strategy for "hard-to-staff" schools. *Review of Educational Research*, *80*(1), 71–107.

Alarcon, G. M. (2011). A meta-analysis of burnout with job demands, resources, and attitudes. *Journal of Vocational Behavior*, *79*(2), 549-562.

Aloe, A. M., Amo, L. C., & Shanahan, M. E. (2014). Classroom management self-efficacy and burnout: A multivariate meta-analysis. *Educational Psychology Review*, *26*(1), 101-126.

*Ancona, M. R., & Mendelson, T. (2014). Feasibility and preliminary outcomes of a yoga and mindfulness intervention for school teachers. *Advances in School Mental Health Promotion*, *7*(3), 156-170.

*Anderson, V. L., Levinson, E. M., Barker, W., & Kiewra, K. R. (1999). The effects of meditation on teacher perceived occupational stress, state and trait anxiety, and burnout. *School Psychology Quarterly*, *14*(1), 3.

*Anderson, S. (2000). *The effects of a rational emotive behavior therapy intervention on irrational beliefs and burnout among middle school teachers in the State of Iowa.* [Doctoral dissertation, University of Northern Iowa]. Ben & Howell Information and Learning Company.

*Anopchand, N. K. (2000). *The impact of expressive writing on prospective teachers' sense of efficacy, stress, burnout, and satisfaction with teaching* [Doctoral dissertation, University of Florida]. ProQuest Information & Learning.

Armon, G., Melamed, S., Shirom, A., & Shapira, I. (2010). Elevated burnout predicts the onset of musculoskeletal pain among apparently healthy employees. *Journal of Occupational Health Psychology*, *15*(4), 399.

Bakker, A. B., & Demerouti, E. (2007). The job demands-resources model: State of the art. *Journal of Managerial Psychology*, *22*(3), 309-328.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191-215.

*Barr, D. J., Boulay, B., Selman, R. L., McCormick, R., Lowenstein, E., Gamse, B., Fine, M. & Leonard, M. B. (2015). A randomized controlled trial of professional development for interdisciplinary civic education: Impacts on humanities teachers and their students. *Teachers College Record*, *117*(2), 1-52.

Becker, B. J. (2000). Multivariate meta-analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), Handbook of applied multivariate statistics and mathematical modeling (pp. 499–525). San Diego: Academic.

Beltman, S., Mansfield, C., & Price, A. (2011). Thriving not just surviving: A review of research on teacher resilience. *Educational Research Review*, *6*(3), 185-207.

Bianchi, R., Boffy, C., Hingray, C., Truchot, D., & Laurent, E. (2013). Comparative symptomatology of burnout and depression. *Journal of Health Psychology*, *18*(6), 782-787.

Blasé, J. J. (1982). A social-psychological grounded theory of teacher stress and burnout. *Educational Administration Quarterly*, *18*(4), 93-113.

Boersma, K., & Linblom, K. (2009). Stability and change in burnout profiles over time: a prospective study in the working population. *Work and Stress, 23,* 264–283.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H. R. (2009). Meta-Regression. In M. Borenstein (Ed.), *Introduction to meta-analysis* (190-202). Chichester: Wiley & Sons.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

*Bradshaw, C. P., Pas, E. T., Goldweber, A., Rosenberg, M. S., & Leaf, P. J. (2012). Integrating school-wide positive behavioral interventions and supports with tier 2 coaching to student support teams: The PBIS plus model. *Advances in School Mental Health Promotion, 5*(3), 177-193.

*Breeman, L. D., van Lier, P. A., Wubbels, T., Verhulst, F. C., van der Ende, J., Maras, A., Struiksma, A.J.C., Hopman, J.A.B, & Tick, N. T. (2016). Effects of the Good Behavior Game on the behavioral, emotional, and social problems of children with psychiatric disorders in special education settings. *Journal of Positive Behavior Interventions, 18*(3), 156-167.

Brewer, E. W., & Shapard, L. (2004). Employee burnout: A meta-analysis of the relationship between age or years of experience. *Human Resource Development Review, 3*(2), 102-123.

Briner, R., & Dewberry, C. (2007). *Staff wellbeing is key to school success.* London: Worklife Support Ltd/Hamilton House.

Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education, 16*(2), 239-253.

Brown, C. (2012). A systematic review of the relationship between self-efficacy and burnout in teachers. *Educational and Child Psychology, 29*(4), 47–63. doi:10.1080/08865710802111972.

Casteel, C. L. (2018). *The Relationship Between Certification Type and Burnout, Perceived Support, Job Satisfaction, and Life Satisfaction in Special Education Teachers* (Unpublished doctoral dissertation). Texas A&M University-Commerce.

*Castillo, R., Fernández-Berrocal, P., & Brackett, M. A. (2013). Enhancing teacher effectiveness in Spain: A pilot study of the RULER approach to social and emotional learning. *Journal of Education and Training Studies*, *1*(2), 263-272.

*Castillo-Gualda, R., García, V., Pena, M., Galán, A., & Brackett, M. A. (2017). Preliminary findings from RULER Approach in Spanish teachers' emotional intelligence and work engagement. *Electronic Journal of Research in Educational Psychology*, *15*(3), 641-664.

Cavanagh, K., Strauss, C., Forder, L., & Jones, F. (2014). Can mindfulness and acceptance be learnt by self-help?: A systematic review and meta-analysis of mindfulness and acceptance-based self-help interventions. *Clinical Psychology Review*, *34*(2), 118-129.

Center for Reviews and Dissemination. *PROSPERO International prospective register of systematic reviews*. Retrieved from https://www.crd.york.ac.uk/PROSPERO/#index.php

Chang, M. L. (2009). An appraisal perspective of teacher burnout: Examining the emotional work of teachers. *Educational psychology review*, *21*(3), 193-218.

*Cheek, J. R., Bradley, L. J., Parr, G., & Lan, W. (2003). Using music therapy techniques to treat teacher burnout. *Journal of Mental Health Counseling*, *25*(3), 204-217.

Cheon, S. H., Reeve, J., Yu, T. H., & Jang, H. R. (2014). The teacher benefits from giving autonomy support during physical education instruction. *Journal of Sport and Exercise Psychology*, *36*(4), 331-346.

Chiesa, A., & Serretti, A. (2009). Mindfulness-based stress reduction for stress management in healthy people: a review and meta-analysis. *The Journal of Alternative and Complementary Medicine*, *15*(5), 593-600.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Collaborative for Academic, Social, and Emotional Learning [CASEL]. (2013). *Effective social and emotional learning programs – Preschool and elementary school edition*. Chicago, IL: Author

Collaborative for Academic, Social, and Emotional Learning [CASEL]. (2015). Effective social and emotional learning programs – Middle and high school edition. Chicago, IL: Author

*Cooley, E., & Yovanoff, P. (1996). Supporting professionals-at-risk: Evaluating interventions to reduce burnout and improve retention of special educators. *Exceptional Children*, *62*(4), 336-355.

Corbin, C. M., Alamos, P., Lowenstein, A. E., Downer, J. T., & Brown, J. L. (2019). The role of teacher-student relationships in predicting teachers' Personal Accomplishment and Emotional Exhaustion. *Journal of School Psychology*, 77, 1-12.

Cowan, J., Goldhaber, D., Hayes, K., & Theobald, R. (2016). Missing elements in the discussion of teacher shortages. *Educational Researcher*, *45*(8), 460-462.

Darling-Hammond, L., & Rothman, R. (Eds.). (2011). Teacher and leader effectiveness in high performing education systems. Washington, D.C./Stanford, CA: Alliance for Excellent Education/Center for Opportunity Policy in Education.

*de Carvalho, J. S., Pinto, A. M., & Marôco, J. (2017). Results of a mindfulness-based social-emotional learning program on portuguese elementary students and teachers: a quasi-experimental study. *Mindfulness*, *8*(2), 337-350.

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian psychology/Psychologie canadienne*, *49*(3), 182.

De Heus, P., & Diekstra, R. F. W. (1999). Do teachers burn out more easily? A comparison of teachers with other social professions on work stress and burnout symptoms. In R. Vandenberghe & A. M. Huberman (Eds.), *Understanding and preventing teacher burnout: A sourcebook of international research and practice* (pp. 269 –284). New York: Cambridge University Press.

Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology*, *86*(3), 499.

Demerouti, E., Bakker, A. B., Vardakou, I., & Kantas, A. (2003). The convergent validity of two burnout instruments: A multitrait-multimethod analysis. *European Journal of Psychological Assessment*, *19*(1), 12.

*Dicke, T., Elling, J., Schmeck, A., & Leutner, D. (2015). Reducing reality shock: The effects of classroom management skills training on beginning teachers. *Teaching and Teacher Education*, *48*, 1-12.

*Domitrovich, C. E., Bradshaw, C. P., Berg, J. K., Pas, E. T., Becker, K. D., Musci, R., Embry, D.D., & Ialongo, N. (2016). How do school-based prevention programs impact teachers? Findings from a randomized trial of an integrated classroom management and social-emotional program. *Prevention Science*, *17*(3), 325-337.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of pediatric psychology*, *34*(9), 917-928.

*Ebert, D. D., Lehr, D., Boß, L., Riper, H., Cuijpers, P., Andersson, G., Thiart, H., Herber, E., & Berking, M. (2014). Efficacy of an internet-based problem-solving training for teachers: results of a randomized controlled trial. *Scandinavian Journal of Work, Environment & Health*, 582-596.

Edmonson, S. L., & Thompson, D. P. (2000). Burnout among Special Educators: A Meta-Analysis. Paper presented at the Annual Meeting of the University Council for Educational Administration. (14th, Albuquerque, NM, November 3-5, 2000).

*Elder, C., Nidich, S., Moriarty, F., & Nidich, R. (2014). Effect of transcendental meditation on employee stress, depression, and burnout: a randomized controlled study. *The Permanente Journal*, *18*(1), 19.

*Emery, D. W. (2011). *Crisis in education: A call to ACT* [Doctoral dissertation, University of Missouri]. Dissertations. 426. https://irl.umsl.edu/dissertation/426

*Erdman, R. (2014). *Increasing Worker Engagement and Alleviating Teacher Burnout Syndrome*. AT Still University of Health Sciences.

Fernet, C., Austin, S., Trépanier, S. G., & Dussault, M. (2013). How do job characteristics contribute to burnout? Exploring the distinct mediating roles of perceived autonomy, competence, and relatedness. *European Journal of Work and Organizational Psychology*, *22*(2), 123-137.

Fernet, C., Guay, F., Senécal, C., & Austin, S. (2012). Predicting intraindividual changes in teacher burnout: The role of perceived school environment and motivational factors. *Teaching and Teacher Education*, *28*(4), 514-525.

*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, 7*(3), 182-195.

*Frank, J. L., Reibel, D., Broderick, P., Cantrell, T., & Metz, S. (2015). The effectiveness of mindfulness-based stress reduction on educator stress and well-being: Results from a pilot study. *Mindfulness, 6*(2), 208-216.

Freudenberger, H. J. (1974). Staff burn-out. *Journal of social issues, 30*(1), 159-165.

Freudenberger, H. J. (1975). The staff burn-out syndrome in alternative institutions. *Psychotherapy: Theory, Research & Practice, 12*(1), 73.

Gavish, B., & Friedman, I. A. (2010). Novice teachers' experience of teaching: A dynamic aspect of burnout. *Social Psychology of Education, 13*(2), 141-167.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional children, 71*(2), 149-164.

Gilmour, A. F., & Wehby, J. H. (in press). The association between teaching students with disabilities and teacher turnover. *Journal of Educational Psychology*. doi:10.1037/edu0000

Giluk, T. L. (2009). Mindfulness, Big Five personality, and affect: A meta-analysis. *Personality and Individual Differences, 47*(8), 805-811.

Goldhaber, D., Krieg, J., Theobald, R., & Brown, N. (2015). Refueling the STEM and special education teacher pipelines. *Phi Delta Kappan, 97*(4), 56-62.

Grayson, J. L., & Alvarez, H. K. (2008). School climate factors relating to teacher burnout: A mediator model. *Teaching and teacher education, 24*(5), 1349-1363.

Hakanen, J. J., Bakker, A. B., & Schaufeli, W. B. (2006). Burnout and work engagement among teachers. *Journal of School Psychology*, *43*(6), 495-513.

Halbesleben, J. R. (2006). Sources of social support and burnout: A meta-analytic test of the conservation of resources model. *Journal of Applied Psychology*, *91*(5), 1134.

Halbesleben, J. R., & Demerouti, E. (2005). The construct validity of an alternative measure of burnout: Investigating the English translation of the Oldenburg Burnout Inventory. *Work & Stress*, *19*(3), 208-220.

Hallman, T., Thomsson, H., Burell, G., Lisspers, J., & Setterlind, S. (2003). Stress, burnout and coping: differences between women with coronary heart disease and healthy matched women. *Journal of Health Psychology*, *8*(4), 433-445.

*Harris, A. R. (2014). *Promoting educators' social-emotional competence, stress management, and wellbeing through a school-based contemplative intervention: An evaluation of efficacy, implementation process, and mechanisms*. [Doctoral dissertation, The Pennsylvania State University]. ProQuest Dissertations and Theses Global.

Harris, A. R., Jennings, P. A., Abenavoli, R. M., Katz, D. A., Greenberg, M. T., & Schussler, D. (2014, March 6-8). A Daily Dose of CALM: Supporting Middle School Educators' Wellbeing and Classroom Functioning through a Brief Stress Reduction Intervention. In R. Abenavoli (Organizer), *Social and Emotional Interventions in Educational Settings Symposium Cultivating Mindfulness among Educators to Support Their Health, Wellbeing, and Efficacy* [Symposium]. Society for Research on Educational Effectiveness, Washington D.C., United States.

Hastings, R. P., & Bham, M. S. (2003). The relationship between student behaviour patterns and teacher burnout. *School Psychology International*, *24*(1), 115-127.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107-128.

Hobfoll, S. (1989). Conservation of resources: A new attempt at conceptualizing stress. *The American Psychologist, 44*, 513−524.

Honkonen, T., Ahola, K., Pertovaara, M., Isometsä, E., Kalimo, R., Nykyri, E., Aromaa, A., & Lönnqvist, J. (2006). The association between burnout and physical illness in the general population—results from the Finnish Health 2000 Study. *Journal of Psychosomatic Research, 61*(1), 59-66.

Hopman, J. A., van Lier, P. A., van der Ende, J., Struiksma, C., Wubbels, T., Verhulst, F. C., ... & Tick, N. T. (2018). Impact of the Good Behavior Game on special education teachers. *Teachers and Teaching*, *24*(4), 350-368.

Iancu, A. E., Rusu, A., Măroiu, C., Păcurar, R., & Maricuțoiu, L. P. (2018). The effectiveness of interventions aimed at reducing teacher burnout: A meta-analysis. Educational Psychology Review, 30(2), 373-396.

Ingersoll, R., & Kralik, J. M. (2004). *The impact of mentoring on teacher retention: What the research says.* Retrieved from http://environmentalconclusions.com/mentor.php

IntHout, J., Ioannidis, J. P., & Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC medical research methodology*, *14*(1), 25.

Jackson, D., White, I. R., & Thompson, S. G. (2010). Extending Dersimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Statistics in Medicine, 29*(12), 1282–1297.

Jennings, P. A. (2011). Promoting teachers' social and emotional competencies to support performance and reduce burnout. In A. Cohan & A. Honigsfeld (Eds.) *Breaking the mold of pre-service and in-service teacher education: Innovative and successful practices for the 21st century*. New York: Rowman and Littlefield.

*Jennings, P. A., Brown, J. L., Frank, J. L., Doyle, S., Oh, Y., Davis, R., Rasheed, D., DeWeese, A., DeMauro, A.A., Heining, C., & Greenberg, M. T. (2017). Impacts of the CARE for Teachers program on teachers' social and emotional competence and classroom interactions. *Journal of Educational Psychology*, *109*(7), 1010.

*Jennings, P. A., Frank, J. L., Snowberg, K. E., Coccia, M. A., & Greenberg, M. T. (2013). Improving classroom learning environments by Cultivating Awareness and Resilience in Education (CARE): Results of a randomized controlled trial. *School Psychology Quarterly*, *28*(4), 374.

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.

*Johnson, S. M., & Naidoo, A. V. (2013). Transpersonal practices as prevention intervention for burnout among HIV/AIDS coordinator teachers. *South African Journal of Psychology*, *43*(1), 59-70.

*Johnson, S., & Naidoo, A. (2017). Can evolutionary insights into the brain's response to threat suggest different group interventions for perceived stress and burnout of teachers in high-risk schools?. *South African Journal of Psychology*, *47*(3), 401-415.

Kabat-Zinn, J. (1990). *Full catastrophe living: How to cope with stress, pain and illness using mindfulness meditation*. Delta.

Khoury, B., Lecomte, T., Gaudiano, B. A., & Paquin, K. (2013). Mindfulness interventions for psychosis: a meta-analysis. *Schizophrenia Research*, *150*(1), 176-184.

Kokkinos, C. M. (2006). Factor structure and psychometric properties of the Maslach Burnout Inventory-Educators Survey among elementary and secondary school teachers in Cyprus. *Stress and Health*, *22*(1), 25-33.

Kokkinos, C. M. (2007). Job stressors, personality and burnout in primary school teachers. *British Journal of Educational Psychology*, *77*(1), 229-243.

Kristensen, T. S., & Borritz, M. (1999) *Copenhagen Burnout Inventory*. Copenhagen, Denmark: National Institute of Occupational Health.

Kristensen, T. S., Borritz, M., Villadsen, E., & Christensen, K. B. (2005). The Copenhagen Burnout Inventory: A new tool for the assessment of burnout. *Work & Stress*, *19*(3), 192-207.

Kushnir, T., & Melamed, S. (1992). The Gulf War and its impact on burnout and well-being of working civilians. *Psychological Medicine*, *22*(4), 987-995.

Langan, D., Higgins, J. P., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., Viechbauer, W., & Simmonds, M. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research synthesis methods*, *10*(1), 83-98.

*Lantieri, L., & Malkmus, C. D. (2011). *Building inner resilience in teachers and their students*. New York: Metis Associates Evaluation.

Lavian, R. H. (2012). The impact of organizational climate on burnout among homeroom teachers and special education teachers (full classes/individual pupils) in mainstream schools. *Teachers and Teaching*, *18*(2), 233-247.

Leonardo, Z. (2004). The color of supremacy: Beyond the discourse of 'white privilege'. *Educational Philosophy and Theory*, *36*(2), 137-152.

Lipsey, M.D, & Wilson, D. B. (2001). *Practical meta-analysis* (Vol. 49). Thousand Oaks, CA: Sage publications.

Loeb, S., Darling-Hammond, L., & Luczak, J. (2005). How teaching conditions predict teacher turnover. *Peabody Journal of Education*, *80*(3), 44–70.

Lundgren-Nilsson, Å., Jonsdottir, I. H., Pallant, J., & Ahlborg, G. (2012). Internal construct validity of the Shirom-Melamed Burnout questionnaire (SMBQ). *BMC Public Health*, *12*(1), 1.

Maslach, C. (1978). Job burnout: How people cope. *Public Welfare*, *36*(2), 56-58.

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Organizational Behavior*, *2*(2), 99-113.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). Maslach burnout inventory manual. Palo Alto, Calif: Consulting Psychologists Press.

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology*, *52*(1), 397-422.

McCarthy, C. J., Lambert, R. G., Lineback, S., Fitchett, P., & Baddouh, P. G. (2016). Assessing Teacher Appraisals and Stress in the Classroom: Review of the Classroom Appraisal of Resources and Demands. *Educational Psychology Review*, *28*(3), 577-603.

McCarthy, C. J., Lambert, R. G., O'Donnell, M., & Melendres, L. T. (2009). The relation of elementary teachers' experience, stress, and coping resources to burnout symptoms.

McLean, D., Eklund, K., Kilgus, S. P., & Burns, M. K. (2019). Influence of teacher burnout and self-efficacy on teacher-related variance in social-emotional and behavioral screening scores. *School Psychology*, *34*(5), 503.

Melamed, S., Kushnir, T., & Shirom, A. (1992). Burnout and risk factors for cardiovascular diseases. *Behavioral Medicine*, *18*(2), 53-60.

Melamed, S., Shirom, A., Toker, S., Berliner, S., & Shapira, I. (2006). Burnout and risk of cardiovascular disease: evidence, possible causal paths, and promising research directions. *Psychological Bulletin*, *132*(3), 327.

Menselson, T., Rehkopf, D. H., & Kubzansky, L. D. (2008). Depression among Latinos in the United States: a meta-analytic review. *Journal of Consulting and Clinical Psychology*, *76*(3), 355.

Möbius, T. W. D. (2014). Metagen: inference in meta-analysis and meta regression. R Package Version, 1(0).

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med, 6*, 1-6. doi: 10.1371/journal.pmed1000097

Montgomery, C., & Rupp, A. A. (2005). A meta-analysis for exploring the diverse causes and effects of stress in teachers. *Canadian Journal of Education/Revue Canadienne de l'éducation*, 458-486.

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210.

Neves de Jesus, S., & Conboy, J. (2001). A stress management course to prevent teacher distress. *International Journal of Educational Management*, *15*(3), 131-137.

Ng, T. W., & Feldman, D. C. (2012). Employee voice behavior: A meta-analytic test of the conservation of resources framework. *Journal of Organizational Behavior*, *33*(2), 216-234.

Pas, E. T., Bradshaw, C. P., Hershfeldt, P. A., & Leaf, P. J. (2010). A multilevel exploration of the influence of teacher efficacy and burnout on response to student problem behavior and school-based service use. *School Psychology Quarterly*, *25*(1), 13.

Pinquart, M., & Sörensen, S. (2005). Ethnic differences in stressors, resources, and psychological outcomes of family caregiving: A meta-analysis. *The Gerontologist*, *45*(1), 90-106.

*Porter, S. A. (2000). Teacher burnout: A quasiexperimental study of psychoeducational group interventions (Unpublished doctoral dissertation). The University of Wyoming.

Purvanova, R. K., & Muros, J. P. (2010). Gender differences in burnout: A meta-analysis. *Journal of Vocational Behavior*, *77*(2), 168-185.

Raedeke, T.D., & Smith, A.L. (2001). Development and preliminary validation of an athlete burnout measure. *Journal of Sport & Exercise Psychology*, *23,* 281-306.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reichl, C., Leiter, M. P., & Spinath, F. M. (2014). Work–nonwork conflict and burnout: A meta-analysis. *Human Relations*, *67*(8), 979-1005.

Robertson, J. S., & Singleton, J. D. (2010). Comparison of traditional versus alternative preparation of special education teachers. *Teacher Education and Special Education, 33*, 213-224.

*Roeser, R. W., Schonert-Reichl, K. A., Jha, A., Cullen, M., Wallace, L., Wilensky, R., Oberle, E., Thomson, K., Taylor, C., & Harrison, J. (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.

Roffey, S. (2012). Pupil wellbeing—Teacher wellbeing: Two sides of the same coin?. *Educational and Child Psychology*, *29*(4), 8.

RStudio Team (2019). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Ruble, L., & McGrew, J. H. (2013). Teacher and child predictors of achieving IEP goals of children with autism. *Journal of Autism and Developmental Disorders*, *43*(12), 2748-2763.

Salmela-Aro, K., Kiuru, N., Leskinen, E., & Nurmi, J. E. (2009). School burnout inventory (SBI) reliability and validity. *European Journal of Psychological Assessment*, *25*(1), 48-57.

Seidman, S. A., & Zager, J. (1986). The Teacher Burnout Scale. *Educational Research Quarterly*. *11*(1), 26–33.

Schonfeld, I. S., Verkuilen, J., & Bianchi, R. (2019a). An exploratory structural equation modeling bi-factor analytic approach to uncovering what burnout, depression, and anxiety scales measure. *Psychological assessment, 31*(8), 1073–1079.

Schonfeld, I. S., Verkuilen, J., & Bianchi, R. (2019b). Inquiry into the correlation between burnout and depression. *Journal of occupational health psychology*.

Schwarzer, R., & Hallum, S. (2008). Perceived teacher self-efficacy as a predictor of job stress and burnout: Mediation analyses. *Applied Psychology*, *57*(s1), 152-171.

Shin, H., Park, Y. M., Ying, J. Y., Kim, B., Noh, H., & Lee, S. M. (2014). Relationships between coping strategies and burnout symptoms: A meta-analytic approach. *Professional Psychology: Research and Practice, 45*, 44-56.

Shoji, K., Cieslak, R., Smoktunowicz, E., Rogala, A., Benight, C. C., & Luszczynska, A. (2016). Associations between job burnout and self-efficacy: a meta-analysis. *Anxiety, Stress, & Coping*, *29*(4), 367-386.

*Siu, O. L., Cooper, C. L., & Phillips, D. R. (2014). Intervention studies on enhancing work well-being, reducing burnout, and improving recovery experiences among Hong Kong health care workers and teachers. *International Journal of Stress Management*, *21*(1), 69.

Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology*, *99*, 611–625. doi:10.1037/0022-0663.99.3.611.

Sutcher, L., Darling-Hammond, L., & Carver-Thomas, D. (2016). A Coming Crisis in Teaching?. *Teacher Supply, Demand, and Shortages in the US*. Palo Alto, CA: Learning Policy Institute.

Swider, B. W., & Zimmerman, R. D. (2010). Born to burnout: A meta-analytic path model of personality, job burnout, and work outcomes. *Journal of Vocational Behavior*, *76*(3), 487-506.

Tarricone, I., Stivanello, E., Poggi, F., Castorini, V., Marseglia, M. V., Fantini, M. P., & Berardi, D. (2012). Ethnic variation in the prevalence of depression and anxiety in primary care: a systematic review and meta-analysis. *Psychiatry Research*, *195*(3), 91-106.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. Educational Researcher, 31(3), 24-31.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*, 783–805.

Ullrich, A., Lambert, R. G., &McCarthy, C. J. (2012). Relationship of German elementary teachers' occupational experience, stress, and coping resources to burnout symptoms. *International Journal of Stress Management, 19,* 333–342.

*Unterbrink, T., Pfeifer, R., Krippeit, L., Zimmermann, L., Rose, U., Joos, A., Hartman, A., Wirsching, M., & Bauer, J. (2012). Burnout and effort–reward imbalance improvement for teachers by a manual-based group program. *International archives of occupational and environmental health*, *85*(6), 667-674.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1-48. Retrieved from http://www.jstatsoft.org/article/view/v036i03

Villegas, A. M., Strom, K., & Lucas, T. (2012). Closing the racial/ethnic gap between students of color and their teachers: An elusive goal. *Equity & Excellence in Education*, *45*(2), 283-301.

Virgili, M. (2015). Mindfulness-based interventions reduce psychological distress in working adults: A meta-analysis of intervention studies. *Mindfulness*, *6*(2), 326-337.

Wachs, K., & Cordova, J. V. (2007). Mindful relating: Exploring mindfulness and emotion repertoires in intimate relationships. *Journal of Marital and Family Therapy*, *33*(4), 464-481.

West, C. P., Dyrbye, L. N., Erwin, P. J., & Shanafelt, T. D. (2016). Interventions to prevent and reduce physician burnout: a systematic review and meta-analysis. *The Lancet*, *388*(10057), 2272-2281.

Wolf, S., Torrente, C., Frisoli, P., Weisenhorn, N., Shivshanker, A., Annan, J., & Aber, J. L. (2015). Preliminary impacts of the "Learning to Read in a Healing Classroom" intervention on teacher well-being in the Democratic Republic of the Congo. *Teaching and Teacher Education*, *52*, 24-36.

Yu, X., Wang, P., Zhai, X., Dai, H., & Yang, Q. (2015). The effect of work stress on job burnout among teachers: The mediating role of self-efficacy. *Social Indicators Research*, *122*(3), 701-708.

Zarate, K., Maggin, D. M., & Passmore, A. (2019). Meta-analysis of mindfulness training on teacher well-being. *Psychology in the Schools, 56*, 1700-1715.

APPENDIX A

CODING SHEET

Report Characteristics
1) Enter study ID (First 6 letters of first author's last name_ Date of publication, e.g. ANCONA_2014) 2) APA Citation 3) Obtained from? 4) Journal (if applicable)
5) What type of report is this? 1 = Journal Article 2 = Dissertation 3 = MA Thesis 4 = Conference Paper 5 = Other (specify) _____
6) Is this a peer reviewed document? 0 = Not peer-reviewed 1 = Peer-reviewed
Inclusion Criteria
7) Is one target population "elementary school teachers" OR "middle school teachers" OR "secondary school teachers" OR "teacher*" OR "educator*" for K-12 or equivalent age groups (adolescents in vocational school, etc.) 0 = No (if participants are (only) preschool teachers, university faculty, or adult vocational educators, then they do not meet criteria) 1 = Elementary school teachers 2 = Middle school teachers 3 = Secondary school teachers 4 = Teachers or Educators (not specified or combined groups)
8) Study Design (any intervention study should fall in one of these) 0 = Qualitative Design or No Comparison Condition 1 = Experimental Design (clearly specifies random assignment) 2 = Quasiexperimental Design (comparison condition not randomly assigned) 3 = Mixed Methods
9) Did the study include teacher burnout factors (e.g. total burnout or "Emotional Exhaustion" scale) as an outcome? 0 = No 1 = Yes; study included measure(s) of teacher burnout
10) Are means (sd) available for baseline and outcome or for difference between baseline and outcome for at least one time point? 0=No 1=Yes
Checkpoint: if any of items 7-9 are 0, stop coding – study is excluded. If 10 is 0, contact study author and request the baseline and outcome data for the teachers – hold off on coding until received

Participants and Sample Characteristics
11) Participants in study were 0 = Preservice Teachers/ Interns/ Pre-licensure beginning teachers 0.5 = New/ beginning Teachers (less than 2 years) 1 = Current/ established Teachers 2 = Both pre-service and current teachers 3 = Both teachers and non-teaching staff who work in schools
12) How were districts recruited? 9999 = Not specified 1= Convenience sample (chose district area to invite) 2= Random sampling (sent invitations to randomly chosen districts/schools) Other, specify: _____
13) How was assignment to intervention or control groups made? 0= convenience 1= random assignment by district 2 = random assignment by school 3= randomly assigned by teacher/classroom 4= matched for some characteristic 5= other (specify) 9999 =not reported
14) Was there any indication of prior intervention to prevent or treat burnout for the sample of teachers included? 0 = no 1= yes
Number of teacher participants enrolled: 15) Total (Intervention + Control) = _____ 16) Control group _____ 17) Intervention group _____
18) What grade(s) were the teachers teaching? 9999 = not specified 1 = K-1 2 = 2-3 3 = 4-5 4 = Middle School (6, 7, 8) 5 = High School (9, 10) 6 = High School (11, 12) 7= Combination K-12
What was the mean age (SD) of teachers in this study, overall? (enter 9999 if not reported) 19) Total Mean _____ 20) Total (SD) _____ What was the mean age (SD) of teachers in the control group? 21) Control Mean _____ 22) Control (SD) _____ What was the mean age (SD) of teachers in the treatment group? 23) Treatment Mean _____

24) Treatment (SD) _____
25) Did the control and treatment groups differ significantly for age?
    9999 = not reported
    0 = no
    1 = yes

For each column, provide the percent of participants who identified as each race/ethnicity: (9999= not reported)

26) Omitted (% of sample that skipped item)
27) Black/African American
28) Hispanic/Latinx
29) White/Caucasian
30) Asian/Asian American
31) other (specify)
32) Did the control and treatment groups differ by race/ethnicity?
    9999 = not reported
    0 = no
    1 = yes

What was the mean "years of experience" (SD) of teachers in this study, overall?

33) Total Mean _____
34) Total (SD) _____

What was the mean "years of experience" (SD) of teachers in the control group?

35) Control Mean _____
36) Control (SD) _____

What was the mean "years of experience" (SD) of teachers in the treatment group?

37) Treatment Mean _____
38) Treatment (SD) _____
39) Did the control and treatment groups differ significantly for "years of experience"?
    9999= not reported
    0 = no
    1 = yes

Educational status of teachers – indicate percent (if none, leave blank):

40) Bachelor's degree _____
41) Advanced degree (Masters, Doctorate) _____
42) Unknown or other _____
43) Did treatment and intervention groups differ significantly on educational status?
    9999= not reported
    0 = no
    1 = yes

Sex /gender of participants – indicate percent (Not reported =9999)

44) Female
45) Male
46) Transgender/non-binary/gender neutral (specify)
47) Omitted item (percent of participants who did not report gender even though it was asked)
48) Did treatment and intervention groups differ significantly on sex?
    9999= not reported

0 = no
1 = yes

Certification type: Indicate percentage if provided (9999=Not reported)

49) Traditional certification (____)
50) Alternative certification (____)

51) Certifications represented included…
1= Traditional only
2= Alternative only
3= Both traditional and alternative certifications represented
9999= Not reported

52) Did treatment and intervention groups differ significantly on certification type?
9999= not reported
0 = no
1 = yes

What was the percent attrition?

53) Overall?___
9999 = not specified
54) For the control group? ____
9999 = not specified
55) For the treatment group? ____
9999= not specified
56) Were differences in attrition found to be "statistically significant"?
9999 = not specified/ not analyzed
1= no
2= yes

Classroom setting – provide % if not reported - enter 9999

57) General Education _____
58) Special Education _____
59) Bilingual Education _____
60) Other (& specify) _____
61) Did the control and treatment groups vary by classroom setting?
9999 = not reported
1 = no
2 = yes

Setting Characteristics

62) Where was study conducted (country)?
9999 = Not reported
1= United States
2= Canada
3 = Other (specify: _____)

63) Specify "other" country if not US or Canada

64) For US., what state or region was study conducted in? _____
9999 = Not reported /not applicable

65) For studies from the US, what county was study conducted in? _____
9999 = Not reported/not applicable

66) For studies from the US—Use the 2013 NCHS scheme to determine: what is the classification for this county? (fill in the blank)

9999=not reported/not applicable

67) Did the study indicate the type of community?
9999 = Not reported
1= Urban
2= Suburban
3 = Rural
4 = Metropolitan
5 = Other: _____
6 = Combination of 2 or more

68) If previous item is "other", what type of study was indicated?_____

69) What type of school setting?
9999 = not specified
1= public school
2= charter school
3= private school
4= alternative placement (discipline; DAEP)
5= alternative placement for developmental disabilities
6 = alternative placement for emotional/behavioral disabilities
7= combination (specify: _____)

70) Specify if combination/multiple (coded 7 for previous item)

71) Number of schools total _____

Student : Teacher ratio: Insert the proportion (1: #students in classroom);
9999= not reported
72) Total sample _____
73) Control sample _____
74) Treatment sample _____

75) Indicators of SES for schools
    9999= Not mentioned or reported
    1= Statement that schools were majority or mostly received free or reduce lunch or considered high poverty
    2= Statement that schools were primarily high SES/ upper middle class
76) Study reported percent of students reporting free/reduced lunch (indicate %): _____

77) Were teachers compensated for participation?
    0 = no
    1= yes
    9999 = not reported

Intervention Characteristics

78) What is the intervention implemented in the study? _____

79) Was "mindfulness" or "cultivating awareness" included as a treatment approach?
    0 = no
    1= yes
80) Was "meditation" included as a treatment approach?
    0 = no
    1= yes
81) Was "yoga" included as a treatment approach?
    0 = no
    1= yes
82) Was the intervention a specific curriculum for teachers to implement?
    0 = no
    1= yes
83) Was the intervention targeted at students (PBIS, SEL, or Mindfulness for students with teacher training)?
    0 = no
    1= yes
84) Was "homework" (directions to meditate at home, do yoga outside of prescribed sessions, and so on) provided as a treatment approach?
    0= No
    1=Yes

85) The intervention is implemented by and for whom?
    1= Teacher implements intervention on their own (e.g. mindfulness for teachers)
    2= Teacher implementing with students AND on their own (e.g. mindfulness for students and teachers)
    3=Teacher implementing ONLY with students (e.g. SEL, PBISplus)

86) Who was the interventionist for the teacher? (i.e. Who trained the teacher on the intervention?) Write in_____(e.g. graduate research student, certified CARE trainer)
    9999 = not clear/ not specified

87) Was the trainer/interventionist involved in any direct contact with the participants/teachers? (workshops, trainings, meetings)
    0 = no
    1= yes
    9999 = not reported
88) How many total minutes of direct contact did teachers have with trainers/interventionists?
    Provide direct contact in minutes (Total) ____
    0 = no direct contact
    9999 = not specified

89) Over how many weeks did direct contact with trainers/interventionists and teachers take place? (if provided in months, multiply months by 4.5)
    Number in weeks (even if contact did not occur weekly) ____
    9999= not specified
90) What was length of initial training in days?
    (enter 9999 if not provided; 4 hours = .5 days)
91) On how many occasions throughout the intervention did teachers have contact with a trainer or treatment provider?
    Number of times participants had contact (e.g. if weekly for 7 weeks, code 7)
    0 = no direct contact after initial training
    9999 = not specified
92) What type of other training, peer support, or administrator support was provided to the teachers? _____

93) Over how long was the intervention in place for these teachers? (Specify in weeks—if listed in months, multiply months by 4.5)

94) Did the control group receive any kind of homework (directions to meditate at home, directions for bibliotherapy, and so on) or other intervention **at the same time as the treatment group**?
    0= No business as usual or waitlist control group
    1= Yes, there was a placebo/ other intervention provided to control group (eLearning, etc.)
    9999= Not specified

95) Specify intervention received by active control group if any indicated in previous item:
    _____

96) Specify minutes of homework assigned weekly to teachers (homework is defined as directions to meditate at home, do yoga outside of prescribed sessions, or complete online discussion posts, and so on—it does not include implementing SEL, GBG, or other curricula with students)
    Minutes/week: _____
    0 = no "homework" was assigned
    1= "homework" assigned but time not specified

97) How many days per week was "homework" assigned to teachers (does not include implementing SEL or other curriculum with students)?
    Days/week: _____

0 = no "homework" was assigned
1= "homework" assigned but days not specified

98) Over how many weeks was "homework" assigned to the teacher (does not include implementing SEL)?
Weeks total: _____
0 = no "homework" was assigned
1= "homework" assigned but weeks not specified

Fidelity

99) Was treatment fidelity data collected for teacher "homework" completion?
0= No
1= Yes
9999 = Not specified or vague statement of monitoring

100) How was treatment fidelity determined for teacher "homework" completion? _____

101) What was rate/percent fidelity for teacher "homework" completion? _____ (enter 9999 if there was homework and fidelity rate was not specified)

102) Was a treatment fidelity (other than for "homework") approach specified?
0 = no
1= yes

How was treatment fidelity data collected? 0=not present, 1= present

103) Observer coding with fidelity checklist
104) Interventionist completed fidelity checklist during intervention
105) Interventionist completed fidelity checklist after intervention
106) Teacher participants (if not interventionist) reported intervention occurred
107) Specify other treatment integrity method_____

108) What was rate/percent fidelity for teacher homework completion? _____ (enter 9999 if not specified)

109) Are there any other treatment approaches that deserve their own code? (List any other approaches or factors noted)

Outcome measure

110) What burnout measure was used in this study?

For each outcome measure, report if the internal consistency (for current study sample) was greater than or equal to .70

111) Emotional exhaustion PRE-intervention (or use this column to code if only one alpha is provided and the study does not specify if pre or post)
0= < .70
1= > or = .70
9999= not reported

112) Emotional exhaustion POST-intervention
0= < .70
1= > or = .70
9999= not reported

113) Cynicism/Cynicism/Depersonalization PRE-intervention (or not specified if pre or post)
0= < .70
1= > or = .70
9999= not reported

114) Cynicism/Cynicism/Depersonalization POST-intervention
0= < .70
1= > or = .70
9999= not reported

115) Personal accomplishment PRE-intervention (or not specified if pre or post)
0= < .70
1= > or = .70
9999= not reported

116) Personal accomplishment POST-intervention
0= < .70
1= > or = .70
9999= not reported

117) Total or other score PRE-intervention (or not specified if pre or post)
0= < .70
1= > or = .70
9999= not reported

118) Total or other score POST-intervention
0= < .70
1= > or = .70
9999= not reported

119) Was there a statement that previous research has found the burnout measure to yield reliable data (with citation)?
0=No
1=Yes

120) Was there a statement that previous research has found the burnout measure yields valid data (with citation)?
0=No
1=Yes

121) Other measure of teacher emotional stability? (list)

Which variables were included in the outcome measure(s)? 0=No, 1= Yes for each
122) Emotional Exhaustion
123) Cynicism/Depersonalization
124) Personal Accomplishment

125) Total score or Other single score (code "work burnout" on CBI)

126) Was follow-up data collected for BOTH treatment AND Control Groups?
0=No
1=Yes

Results

127) Control Sample size
128) Treatment Sample size

For Emotional Exhaustion

129) Measure derived from: _____
130) Control Mean baseline for Emotional Exhaustion
131) Control SD baseline for Emotional Exhaustion
132) Treatment Mean baseline for Emotional Exhaustion
133) Treatment SD baseline for Emotional Exhaustion
134) Did groups differ significantly at baseline for Emotional Exhaustion? (run t-test if necessary)
0=No
1=Yes
135) Control Mean post intervention for Emotional Exhaustion
136) Control SD post intervention for Emotional Exhaustion
137) Treatment Mean post intervention for Emotional Exhaustion
138) Treatment SD post intervention for Emotional Exhaustion
139) ES for Emotional Exhaustion at post-intervention (calculated as Hedges g)
140) ES for Emotional Exhaustion (reported)
141) ES type for Emotional Exhaustion (e.g., Cohen's d)
142) Was there a follow-up conducted?
0=No
1=Yes
143) If yes, how many days after intervention completed? _____
144) What was the Sample Size for the Control Group at Follow-Up?
145) What was the Sample Size for the Treatment Group at Follow-Up?
146) Control Mean follow-up for Emotional Exhaustion
147) Control SD follow-up for Emotional Exhaustion
148) Treatment Mean follow-up for Emotional Exhaustion
149) Treatment SD follow-up for Emotional Exhaustion
150) Calculate ES for Emotional Exhaustion at follow-up as Hedges g
151) ES for Emotional Exhaustion at follow-up (reported)
152) ES type for Emotional Exhaustion (e.g., Cohen's d)

For Cynicism/Depersonalization

153) Measure derived from: _____
154) Control Mean baseline for Cynicism/Depersonalization
155) Control SD baseline for Cynicism/Depersonalization
156) Treatment Mean baseline for Cynicism/Depersonalization
157) Treatment SD baseline for Cynicism/Depersonalization

158) Did groups differ significantly at baseline for Cynicism/Depersonalization? (run t-test to figure out if necessary)
0=No
1=Yes
159) Control Mean post intervention for Cynicism/Depersonalization
160) Control SD post intervention for Cynicism/Depersonalization
161) Treatment Mean post intervention for Cynicism/Depersonalization
162) Treatment SD post intervention for Cynicism/Depersonalization
163) Calculate ES for Cynicism/Depersonalization at post-intervention as Hedges G
164) ES for Cynicism/Depersonalization at post-intervention (reported)
165) ES type for Cynicism/Depersonalization (e.g., Cohen's d)
166) Was there a follow-up conducted?
0=No
1=Yes
167) If yes, how many days after intervention completed? _____
168) Sample size for Control Follow-up
169) Sample size for Treatment Follow-up
170) Control Mean follow-up for Cynicism/Depersonalization
171) Control SD follow-up for Cynicism/Depersonalization
172) Treatment Mean follow-up for Cynicism/Depersonalization
173) Treatment SD follow-up for Cynicism/Depersonalization
174) Calculate ES for Cynicism/Depersonalization at follow-up as Hedges G
175) ES for Cynicism/Depersonalization at follow-up (reported)
176) ES type for Cynicism/Depersonalization (e.g., Cohen's d)

For Personal Accomplishment
177) Measure derived from: _____
178) Control Mean baseline for Personal Accomplishment
179) Control SD baseline for Personal Accomplishment
180) Treatment Mean baseline for Personal Accomplishment
181) Treatment SD baseline for Personal Accomplishment
182) Did groups differ significantly at baseline for Personal Accomplishment?
0=No
1=Yes
183) Control Mean post intervention for Personal Accomplishment
184) Control SD post intervention for Personal Accomplishment
185) Treatment Mean post intervention for Personal Accomplishment
186) Treatment SD post intervention for Personal Accomplishment
187) Calculate ES for Personal Accomplishment post-intervention as Hedges G
188) ES for Personal Accomplishment post-intervention (reported)
189) ES type Personal Accomplishment (e.g., Cohen's d)
190) Was there a follow-up conducted?
0=No
1=Yes
191) If yes, how many days after intervention completed? _____
192) Control Mean follow-up for Personal Accomplishment

193) Control SD follow-up for Personal Accomplishment
194) Treatment Mean follow-up for Personal Accomplishment
195) Treatment SD follow-up for Personal Accomplishment
196) Calculate ES for Personal Accomplishment at follow-up as Hedges G
197) ES for Personal Accomplishment at follow-up (reported)
198) ES type Personal Accomplishment (e.g., Cohen's d)

For Burnout Total Score

199) Measure derived from: _____
200) Control Mean baseline for Total Score
201) Control SD baseline for Total Score
202) Treatment Mean baseline for Total Score
203) Treatment SD baseline for Total Score
204) Did groups differ significantly at baseline for Total Score?
0=No
1=Yes
205) Control Mean post intervention for Total Score
206) Control SD post intervention for Total Score
207) Treatment Mean post intervention for Total Score
208) Treatment SD post intervention for Total Score
209) Calculate ES for Total Score at post-intervention as Hedges g
210) ES for Total Score at post-intervention (reported)
211) ES type for Total Score (e.g., Cohen's d)
212) Was there a follow-up conducted?
0=No
1=Yes
213) If yes, how many days after intervention completed? _____
214) Control Mean follow-up for Total Score
215) Control SD follow-up for Total Score
216) Treatment Mean follow-up for Total Score
217) Treatment SD follow-up for Cynicism/Depersonalization
218) Calculate ES for Total Score at follow-up as Hedges g
219) ES for Total Score at follow-up (reported)
220) ES type for Total Score (e.g., Cohen's d)

221) If social validity/satisfaction survey, when was it administered? (fill in the blank; e.g. immediately after intervention, 6 month follow-up, etc.)

222) What were results of the social validity/satisfaction survey? (fill in the blank, % very satisfied, etc.)

223) What other teacher outcomes were measured? (fill in the blank)

224) Was there a social validity/ satisfaction survey or interview? (include in study quality total score)
0= No

1=Yes
Study Quality (Based on Gersten Standards: Essential Quality Indicators)
Participants
225) Were appropriate procedures used to **increase the likelihood** that relevant characteristics of participants in the sample were comparable across conditions? 0= No (does not specify assignment to groups or is based on convenience, randomly assigned by district) 2 = randomly assigned by school 4 = Yes; random assignment by teacher/class or matching design
226) Was sufficient information given characterizing the interventionists provided? 0= No; interventionists unknown, or may have systematically varied across conditions 1= Yes; interventionists clearly specified (level of training, profession, etc.)
227) Was sufficient demographic information provided (age, sex, race/ethnicity, years' experience, type of certification, educational level) for the teachers? 0= No (demographic data limited to 1-2 of these factors) 1 = Demographic data includes 3+ of these factors
228) Was sufficient information provided or found on the community setting (country, state, county, urban/suburban/rural, SES)? 0= No only vague descriptors of any of these factors (country, state) 1 =Information provided on county, OR type of setting (urban, rural, suburban, residential, etc.) OR on SES of area 2= Information provided both SES AND the type of setting (urban, rural, suburban, residential, etc.)
229) Was the classroom (general ed/special ed, teacher: student ratios) and school (public, private, so on) information for the teachers provided? 0= No only vague descriptors of any of these factors 1 =Information provided on some, but not all factors 2= Information provided on all factors
230) Did teachers in the control and intervention group differ significantly on any of the demographic factors (age, race ethnicity, SES, sex)? 2= No between group differences were present at baseline 1= Minimal between group differences were present or any differences were accounted for, OR article stated that "no demographic differences at baseline" WITHOUT explicitly specifying which demographics were examined. 0= Insufficient information to determine if between group differences existed at baseline
Implementation

231) Was the intervention clearly described and specified? (included information about the duration of intervention, direct contact hours, and what occurred during direct contact hours, if applicable)
0 = No; intervention not clearly described
1 = Yes; intervention clearly described

232) Was the fidelity of implementation described?
0= No; no treatment integrity data were collected
1= Yes; treatment integrity data were collected (for homework/ mindfulness logs only)
2 = Yes; treatment integrity data were collected (not just for homework, but for the intervention itself) and how assessed was specified

233) Is there a record of participants' attendance (or online work completion for online interventions without direct contact)?
0= No; no record of minutes/hours of intervention received by each participant
1= Yes; there is a clear record of how many minutes/hours of intervention each participant received

234) Did the study provide evidence of equivalent burnout scores at baseline?
0= No; baseline differences were statistically significant OR the study did not report any baseline data for burnout.
1= Yes; evidence of equivalent scores at baseline was presented for relevant scales (non-statistically significant differences at p=.05)

Outcome Measures

235) Were **multiple measures** used to provide an appropriate balance between measures closely aligned with the interventions and measures of generalized performance?
0= No; only a burnout measure was used
1= Yes; multiple measures were used

Study Quality 2 (Based on Gersten Standards: Desirable Quality Indicators)

236) Is overall attrition less than 30%?
0= No; overall attrition was **not documented** or was over 30%
1= Yes; overall attrition was less than 30%

237) Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?
0= No; data collectors and/or scorers were not blind to conditions, were not equally familiar to examinees across conditions, or no information was provided on this
1= Yes; data collectors and scorers were blind to condition AND equally familiar/unfamiliar to examinees across conditions

238) Were outcomes for capturing the intervention's effect measured beyond an immediate posttest? (follow-up data collected?)
0= No; no follow-up data
1= Yes; follow-up data were collected for BOTH Treatment AND Control groups

239) Did the research team assess not only surface features of fidelity implementation (e.g., number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine **quality of implementation**?
0= No; quality of implementation was not clearly assessed or documented
1= Yes; quality of implementation was clearly assessed and documented

240) Was any documentation of the nature of instruction or series provided in comparison conditions?
0= No; instruction provided to comparison condition not clearly documented
1= Yes; instruction for comparison condition clearly documented

241) Did the research report include actual audio or videotape excerpts that capture the nature of the intervention?
0= No excerpts included
1= Yes; or participants were instructed to listen to recordings as part of the intervention and excerpts of recording are included

242) Were results presented in a clear, coherent fashion?
0= No; results unclear
1= Yes; results clear

243) Total study quality score
Sum of scores from items 225 through 242, inclusive. First item =224 regarding social validity, Last item = results clear

Coder and Coding Characteristics

244) Who is coding this study? (first 7 letters of last name)

245) Provide any notes or concerns you had about coding this study:

APPENDIX B

SUPPLEMENTARY TABLE FOR STUDY DEMOGRAPHICS

	Total Mean years experience	**Total SD years experience**	**Total Mean Age**	**Total SD Age**	**% Female**	**% Male**	**% non-binary/gender queer**	**% White/ Caucasian**
Ancona & Mendelson, 2014	N/R	N/R	N/R	N/R	81.4%	18.6%	N/R	41.86%
Anderson et al., 1999	N/R	N/R	N/R	N/R	84.62%	15.38%	N/R	93%
Anderson, 2000	14.99	N/R	40.07	N/R	74.8%	25.2%	N/R	N/R
Anopchand, 2000	N/R	N/R	N/R	N/R	88.4%	11.6%	N/R	N/R
Barr et al., 2015	8.49	7.58	N/R	N/R	N/R	N/R	N/R	81%
Bradshaw et al., 2012	7.48	8.35	N/R	N/R	91%	9%	N/R	75.2%
Breeman et al., 2016	14.01	N/R	38.81	N/R	72%	N/R	N/R	N/R
Castillo et al., 2013	15	9.7	44.27	N/R	70%	30%	N/R	N/R
Castillo-Gualda et al., 2017	7.54	4.54	35.98	8.45	79%	N/R	N/R	N/R
Cheek et al., 2003	8.12	6.39	N/R	N/R	94.12%	5.77%	N/R	100%
Cooley, 1995	9.7	6.6	41.6	7.6	89.6%	10.4%	N/R	N/R
de Carvalho et al., 2017	16.44	5.4	40.37	6.3	100%	0%	N/R	N/R
Dicke et al., 2015	N/R	N/R	27.6	4.1	69%	31%	N/R	N/R
Domitrovich et al., 2016	N/R	N/R	N/R	N/R	88%	N/R	N/R	N/R
Ebert et al., 2014	19	9.6	47.1	8.2	83.33%	N/R	N/R	N/R
Elder et al., 2014	N/R	N/R	36.1	N/R	N/R	47.5%	N/R	97.5%
Emery, 2011	17	N/R	47	N/R	82.86%	17.14%	N/R	65.71%
Erdman, 2014	N/R	N/R	N/R	N/R	N/R	N/R	N/R	N/R
Flook et al., 2013	12.83	8.68	43.06	9.87	88.89%	N/R	N/R	94%
Frank et al., 2015	N/R	N/R	40.72	10.77	77.78%	22.22%	N/R	97.22%
Harris et al., 2014	13.96	9.01	42.55	12.53	87.5%	N/R	N/R	98.39%
Jennings et al., 2013	11.7	N/R	36	N/R	89%	11%	N/R	88.68%
Jennings et al., 2017	12.5	N/R	41.5	N/R	93%	7%	N/R	33%
Johnson & Naidoo, 2013	N/R	N/R	N/R	N/R	90.74%	9.26%	N/R	5.56%
Johnson & Naidoo, 2017	N/R	N/R	46	N/R	N/R	N/R	N/R	N/R

Lantieri & Malkmus, 2011	N/R	N/R	N/R	N/R	N/R	N/R	N/R	N/R
Porter, 2000	N/R	N/R	N/R	N/R	N/R	N/R	N/R	N/R
Roeser et al., 2013	14.9	8.5	46.9	9.2	89%	N/R	N/R	79.65%
Siu et al., 2014	13.63	6.91	39.06	8.19	63.27%	36.74%	N/R	N/R
Unterbrink et al., 2012	N/R	N/R	N/R	N/R	N/R	N/R	N/R	N/R
*N/R signifies that data were not reported within the study, thus these are missing data.								

APPENDIX C

STUDIES THAT INCLUDED FOLLOW-UP DATA

Citation	Scales	Latency of Follow-Up	Study Quality Score
Anderson et al., 1999	Emotional Exhaustion Cynicism/ Depersonalization Personal Accomplishment	9 weeks	14
Dicke et al., 2015	Emotional Exhaustion only	10-12 months	6
Ebert et al., 2014	Emotional Exhaustion Cynicism/ Depersonalization Personal Accomplishment	6 months	12
Roeser et al., 2013	Total Burnout Score only	3 months	15