CrossMark

The Effectiveness of Interventions Aimed at Reducing Teacher Burnout: a Meta-Analysis

Alina Eugenia Iancu[1] · **Andrei Rusu**[1] ·
Cristina Măroiu[1] · **Roxana Păcurar**[1] ·
Laurențiu P. Maricuțoiu[1] ⬤ID

Abstract We conducted a meta-analysis to investigate the effectiveness of interventions aimed at reducing teacher burnout. Online and reference list searches yielded 513 unique results, and the final sample contains 23 controlled trials (19 journal articles and 4 unpublished dissertations). More than two thirds of the studies had optimal quality, and the risk of bias was not related to the overall effectiveness of the interventions. The overall effects were small, but statistically significant $(d = 0.18, SE = 0.05, Z = 3.26, p < 0.001, k = 23)$. Separate analyses on each burnout component showed similar intervention effects on emotional exhaustion and personal accomplishment, but almost null effects on depersonalization $(d = 0.03, SE = 0.06, Z = 0.53, p > 0.05, k = 11)$. Additional moderator analyses suggested that mindfulness interventions had significant effects on exhaustion and personal accomplishment. Interventions on primary and middle school teachers reported effect sizes below the average effect. Similar to previous findings, interventions that lasted less than 1 month had the smallest levels of efficacy.

Teacher burnout is becoming a concern for educational institutions, and current research studies suggest that it has serious consequences for the teachers' occupational health and for the educational outcomes of that particular teacher. Regarding the effects on the teacher, previous research revealed that burnout is associated with poor job satisfaction (Domitrovich et al. 2016), high rates of absenteeism (Wolf et al. 2015), anxiety and depression, high blood pressure, or even cardiovascular disease (Roeser et al. 2013). Regarding the educational outcomes, teacher burnout is associated with reduced quality of performance and classroom instruction (Wolf

✉ Laurențiu P. Maricuțoiu
laurentiu.maricutoiu@e-uvt.ro

1   Department of Psychology, West University of Timişoara, 4 Vasile Pârvan Blvd., Room 303,
300223 Timişoara, Romania

⬤ Springer

et al. 2015) and with diminished capacity to engage and effectively teach (Roeser et al. 2013). Furthermore, previous research studies linked teacher stress to poorer classroom climate as well as more unsatisfactory student behavior and achievement (Wolf et al. 2015).

Given these consequences, educational practitioners developed various types of interventions aimed at reducing teacher burnout, as described later in this review. However, we know very little regarding the effectiveness of these interventions. Previous reviews of interventions aimed at reducing burnout showed that interventions have little impact on this phenomenon (see Maricuțoiu et al. 2016 for a general review of interventions; see Panagioti et al. 2016 for a review of interventions on physicians). Regarding teacher burnout, researchers developed controlled trials in high-income countries (e.g., USA—Domitrovich et al. 2016; Flook et al. 2013; Jennings et al. 2013; Germany—Ebert et al. 2014; Unterbrink et al. 2012), and in low-income countries as well (e.g., South Africa—Johnson and Naidoo 2013). However, little is known regarding their overall effectiveness.

In this review, we start from the assumption that the teaching activity has unique occupational stressors (McCarthy et al. 2016); therefore, interventions aimed at reducing burnout should address them in particular. Consequently, we argue that interventions on teachers should be analyzed separately from other occupational categories. Based on these considerations, this review has two major objectives: (a) to conduct a systematic search of the available literature and to (b) analyze the evidence regarding the controlled studies that aimed at reducing teacher burnout.

The Specific Characteristics of Teacher Burnout

Burnout is the response to prolonged exposure to stressors (Maslach et al. 2001), characterized by three components: the lack of resources for handling emotional events (emotional exhaustion), detachment and cynical attitudes towards own job (depersonalization or cynicism), and an intense feeling of professional inefficacy. Because previous reviews reported that interventions do not have the similar effect on the three burnout components (Maricuțoiu et al. 2016), in the present review, we will analyze them separately.

Teacher burnout is directly related to teaching-specific stressors; therefore, previous research studies attempted to clarify the nature of these stressors. Some researchers suggested that the primary stressors of teachers are the socio-emotional demands of working with more than 30 students at once, and the fact that teachers have to make hundreds of decisions "on the fly" each day (Roeser et al. 2012). In a similar vein, Unterbrink et al. (2012) stated that teaching-specific stressors are related to classroom management and include the emotional climate, the dyadic teacher-student relationships, and the interpersonal conflicts with pupils, parents, or colleagues. Other researchers (McCarthy et al. 2016) suggested that teacher burnout results from the unbalance between teaching demands (e.g., problematic student behaviors, administrative demands) and teaching resources (e.g., school support personnel, the existence of instructional materials),

Self-reported data collected from teachers indicated that the main stressors of their job are the workload, the lack of cooperative time with colleagues, the lack of support from superiors, and the management of difficult students in the classroom (Roeser et al. 2013). Additionally, Wolf et al. (2015) noted that teachers in low-income countries face many challenges like increasing workloads due to education reform, low and infrequent teacher remuneration, lack of professional recognition and autonomy, lack of opportunities for professional development,

difficult working conditions, lack of autonomy, and lack of voice. Teachers from South Africa are exposed to additional stressors not commonly encountered in the profession, namely the high prevalence of HIV/AIDS. Hence, volunteer or assigned, they have to take on the role of HIV/AIDS coordination in the schools. Motivated by social caring and helping motives, HIV/AIDS coordinators are at risk of stress and burnout (Johnson and Naidoo 2013).

Therefore, given the specific occupational stressors, we formulated the first research question:

Question1.   *Are interventions effective in reducing teacher burnout?*

Approaches Aimed at Reducing Teacher Burnout

Occupational self-compassion is one of the specific characteristics of teachers which help them to motivate and teach students (Roeser et al. 2013). Cooley and Yovanoff (1996) also included collegial isolation and role conflict or ambiguity as other work-related variables. Additionally, teachers are role models for the kind of skills and mindsets that students in the twenty-first century need in order to be successful (Roeser et al. 2013).

There are several approaches regarding the type of intervention that is needed to decrease teacher burnout. Based on the literature reviewed here, we classified these approaches into the following categories: cognitive behavioral therapy (CBT), mindfulness and relaxation, social-emotional skills, psychoeducational approach, social support, and professional development.

Cognitive Behavioral Therapy

Interventions aimed at enhancing employee coping skills are traditional in occupational health psychology (Maricuțoiu et al. 2016) and involve the use of a cognitive behavioral approach to stress. When used in an educational setting, these approaches provided mixed results. Cooley and Yovanoff (1996) implemented two interventions which consisted of (a) a series of stress management coping skill workshops aimed at preventing or mitigating teacher burnout and (b) a peer collaboration program designed to facilitate supportive, collegial interactions among teachers regarding work-related problems. The treatment group made a desirable change in the case of depersonalization and personal accomplishment, whereas the control group showed undesirable change. The treatment group also showed relatively greater improvement (decrease) in emotional exhaustion (Cooley and Yovanoff 1996). On the other hand, Ebert et al. (2014) reported no between-group differences in change from baseline to post-treatment, but the changes from baseline to 6-month follow-up were only significant for depersonalization and emotional exhaustion. In their study, Ebert et al. (2014) aimed at evaluating the efficacy of Internet-based problem-solving training (iPST) for teachers with a heightened level of depressive symptoms.

Interventions Based on Mindfulness and Relaxation Techniques

Interventions based on mindfulness and relaxation techniques provided encouraging results in previous literature reviews (Maricuțoiu et al. 2016; Richardson and Rothstein 2008). Roeser et al. (2013) considered mindfulness as a useful intervention in diminishing the levels of teacher burnout and identified three change mechanisms that

can explain the mindfulness utility in reducing burnout (Roeser et al. 2013). Firstly, mindfulness develops awareness of the antecedents to one's stress reaction (e.g., what does generates my emotional reactions and how can I use this information in order to reduce stress?). Secondly, it develops awareness of the bodily sensations that accompany being "stressed out" (e.g., knowing what I am feeling). Lastly, it generates a set of strategies for coping effectively with stress (e.g., taking a break and breathing deeply before doing something, escaping ruminative thinking in favor of focusing on present moments, letting go of highly rated expectations and illusions of control, seeing the pain and reason behind others' difficult behavior rather than taking it personally, and being compassionate with oneself when something is wrong). Roeser et al. (2013) found that the effect sizes with regard to reductions in teachers' burnout at post-program and follow-up were large. On the other hand, Flook et al. (2013) adapted their course specifically for teachers, and they identified significant improvements in emotional exhaustion and personal accomplishment components.

Social-Emotional Skills

The quality of the teacher-student relationship is considered an important factor that contributes to teacher well-being (Spilt et al. 2011). The development of social-emotional skills should improve these relationships, which, in turn, could reduce teacher burnout. Social-emotional skills include the development of supportive relationships with students, the management of challenging student behaviors, and to provide modeling and direct instruction for effective social and emotional learning (Jennings et al. 2013). Regarding the results obtained by the authors who integrated into their studies this type of intervention, Jennings et al. (2013) found significant intervention effects only on the personal accomplishment subscale, while Wolf et al. (2015) identified no statistically significant effects on general burnout.

Psychoeducational Approach

The psychoeducational approach aims to increase the teacher knowledge regarding the prevalence of stress and burnout within the education field (Emery 2011). Emery (2011) reported that burnout levels decreased for the experimental group, yet increased for the control group. Unterbrink et al. (2012) mixed the psychoeducational approach with social support groups and reported positive effects of the intervention on emotional exhaustion and personal accomplishment.

Social Support

The fifth type of approach refers to social support. Social support involves the use of group work, in which case the teachers should feel supported and encouraged for their work by their colleagues (Unterbrink et al. 2012). Small to moderate effects were found by Unterbrink et al. (2012) for emotional exhaustion and personal accomplishment. On the other hand, Cooley and Yovanoff (1996) showed that the treatment group made a desirable change in the case of depersonalization and personal accomplishment, and relatively greater improvement in emotional exhaustion, whereas the control group showed undesirable change.

Professional Development

Professional development strategy specifies the fact that teachers are trained through didactic lessons to provide explicit instruction to students to promote the development of emotional awareness and communication, self-regulation, social problem solving, and relationship management skills (Berg et al. 2016). Breeman et al. (2016) found no impact of the intervention over emotional exhaustion and personal accomplishment. On the other hand, Cheon et al. (2014) showed that emotional-physical exhaustion decreased significantly for teachers in the experimental group, while it remained unchanged for teachers in the control group.

Given these different approaches present in the literature, we formulated our second research question as follows:

Question2. *Are all intervention types equally effective in reducing teacher burnout?*

Other Potential Moderator Variables

Teaching Level

The main levels at which an educator can teach are (1) primary level (in which children receive primary or elementary education from the age of about 5 to 12), (2) middle level (which consists of students with ages that vary from about 12 to 15), and (3) high school (which is comprised by adolescents that prepare themselves for future jobs). These levels of education require different educational activities and different student-teacher relationships. For example, Hargreaves (2000) reported that elementary school teachers reported more intense emotions in the classroom, as compared with secondary school teachers. This finding suggests that elementary school teachers can have different demands, as compared to secondary school teachers. Therefore, similar interventions can have considerably different efficacy from one level to another. Based on these ideas, we formulated our third research question as follows:

Question3. *Are interventions equally effective at different teaching levels?*

Time Lag

The time lag between the end of the intervention and the assessment of intervention efficacy is highly debated in the previous literature, which is why we chose to integrate it in our review as well. Specifically, Maricuțoiu et al. (2016) showed that for emotional exhaustion, the intervention effectiveness remains the same at different time points, and in the case of depersonalization and personal accomplishment, the intervention effectiveness is zero regardless the moments of assessment. However, Maricuțoiu et al. (2016) also reported that post-intervention effects are highly heterogeneous; therefore, it is possible to find stronger effect sizes in post-intervention assessments, as compared with follow-up assessments. Therefore, we formulated our fourth research question as follows:

Question4. *Does intervention efficacy vary as a function of the time lag between the end of intervention and the assessment moment?*

Intervention Duration

The length of the intervention is also highly debated in previous occupational health reviews. For example, Richardson and Rothstein (2008) reported similar effect sizes for short interventions (less than 4 weeks) and for medium-length interventions (5–12 weeks). On the other hand, Maricuțoiu et al. (2016) also analyzed this moderator and reported null effects of interventions that last less than a month, and stronger effects of lengthier interventions. Starting from these divergent findings, we formulated our last research question as follows:

Question5.   *Does intervention efficacy vary as a function of the intervention length?*

Method

Literature Search

The final search was conducted during February 2017 through the EBSCOhost interface and interrogated the following databases: Academic Search Complete, Academic Search Premier, Central and Eastern European Academic Source, EconLit, Education Research Complete, MEDLINE, Middle Eastern and Central Asian Studies, Psychology and Behavioral Sciences Collection, PsycINFO, and Teacher Reference Center. We used this exact query of keywords connected with Boolean operators: (burnout OR exhaustion OR cynicism OR depersonalization OR inefficacy OR "personal accomplishment") AND (teacher OR educator OR instructor OR professor) AND (trial OR intervention). No other search restraints were imposed. Additionally, we looked for possibly eligible articles in the reference lists of the existing systematic or qualitative reviews on burnout.

Inclusion/Eligibility Criteria

The eligibility criteria for the included studies were (1) the research had to assess general burnout, core burnout, or its components (any form of exhaustion, depersonalization, or cynicism and personal accomplishment); (2) the research had to assess these variables both at pre-test and post-test; (3) the design had to include a passive control group (i.e., no intervention or a waiting list group); (4) the target group had to be comprised of teachers (of any level or type of education); (5) the burnout levels of the compared groups had to be equivalent at the baseline moment (i.e., randomized allocation into groups and/or no statistical difference between groups in terms of targeted outcomes at baseline); (6) and the authors had to report statistical indices (e.g., means and standard deviations, *t* tests) in order to compute the effect size regarding the differences between experimental and control groups.

Study Coding and Quality Assessment

We carried out the study coding in two stages. Firstly, we independently analyzed each eligible article and selected the relevant information for computing effect sizes (the sample size of the experimental and control group, and means and standard deviations for pre-intervention and all

post-intervention assessments). Also at this stage, we coded the time lag between the end of the intervention and the outcome assessment. In order to establish a higher accuracy of the data collected in this stage, the information was compared and verified by the two authors.

Secondly, we selected the following studies' characteristics: identification data (author(s) and year of publication), the level of teaching education (e.g., elementary, middle, or high level of education), the nationality of the participants, intervention length, a short description of the intervention, and the intervention approach. Although we did not impose any restrictions regarding the language of the papers, all interventions included in this review are written in English. All of these characteristics are described in detail in Table 1.

Study quality (risk of internal bias) was assessed based on the Cochrane Collaboration tool (Higgins and Green 2011). Two independent raters scrutinized each study regarding the six criteria proposed in the tool (i.e., sequence generation, allocation concealment, blinding of outcome assessor, incomplete outcome data, selective outcome reporting, other potential threats to validity). For each criterion, the raters assessed whether the study is at *low risk of bias*, *high risk of bias*, or *the risk of bias is unclear*. For each study, we computed three scores representing the total number of criteria on which it was classified as having a low, unclear, or high risk of bias (see Table 1; last three columns on the right side). Moreover, for each criterion, we computed the percentage of studies which were classified as having low, unclear, or high risk of bias (see Fig. 4).

In order to estimate the degree of agreement between the independent assessors, we calculated the kappa statistic and interpreted it based on the following benchmark (Gwet 2012): 0.00–0.20 = slight agreement, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, and 0.81–1.00 = almost perfect agreement. The results ranged between fair (kappa = 0.32; for blinding of outcome assessors) to substantial (kappa = 0.71; for selective outcome reporting), with a median agreement of 0.47 for sequence generation (selection bias). Any incongruence between experts was consensually solved with the mediation of the corresponding author.

Meta-Analytical Procedure

The meta-analysis was conducted with the aid of Comprehensive Meta-Analysis version 2.0 (Borenstein et al. 2005). Due to the mixed characteristics of the studies, we assumed a random effects model. We took into account the following indicators: k (the number of studies included in the meta-analysis), d (the average effect size expressed in Cohen's index, with values around 0.20 indicating small effect sizes), SE (standard error of the average effect size), lower limit and upper limit effect sizes (the values of the 95% confidence interval), Z (statistical test for the null hypothesis regarding the average effect), and the indicators of heterogeneity, namely Q and I^2. The Q test is used for testing whether the differences between the studies and their averaged effect are either marginally or statistically significant. The I^2 index estimates the percentage of effect variance that can be attributed to systematic variations between the studies, and values between 0.25 and 0.50 indicate acceptable proportions of between-study variance attributable to moderator variables (Borenstein et al. 2009).

Because studies reported multiple effect sizes based on the same participants (e.g., effect sizes from post-intervention and follow-up), we used the algorithms developed by Borenstein et al. (2009, p. 225–238), and all these algorithms are implemented by the Comprehensive Meta-Analysis software (Borenstein et al. 2005). We used the study as the unit of analysis, and

Table 1 Overview of the studies included in the meta-analysis

Study	School level	Country	Outcome	Intervention length	Approach							Risk of bias		
					Cognitive behavioral	Mindfulness/ meditation	Professional development	Psycho educational	Social support	Socio-emotional skills	Other	Low risk	Unclear	High
Ancona and Mendelson (2014)	Elementary and middle school	USA	Emotional exhaustion (MBI-ES; Maslach et al. 1996)	3 weeks		x						2	3	1
Anderson (2000)	Middle school	USA	Emotional exhaustion, depersonalization, personal accomplishment (MBI-ES; Maslach et al. 1986)	–	x			x				4	1	1
Anderson et al. (1999)	Elementary, middle, and high school	USA	Emotional exhaustion, depersonalization, personal accomplishment (MBI-ES; Maslach and Jackson 1993)	5 weeks		x						3	3	0
Anopchand (2000)	Elementary school	USA	Emotional exhaustion (ES; Maslach et al. 1986)	One/two sessions of 20 min each							Expressive writing	5	1	0
Berg et al. (2016)	Elementary school	USA	Emotional exhaustion, depersonalization, personal accomplishment (MBI; Maslach et al. 1996)	One school year			x					2	2	2
Berg et al. (2016)	Special primary education	Netherlands	Emotional exhaustion, personal accomplishment (UBOS-L; Schaufeli and Van Dierendonck 2000)	One school year			x					2	2	2
Cheon et al. (2014)	Elementary, middle, and high school	South Korea	Emotional-physical exhaustion (ABQ; Stebbings et al. 2012)	Fall semester (approx. 4 months)			x					5	1	0
Cooley and Yovanoff (1996)	Special education (elementary, middle, and high)	USA	Emotional exhaustion, depersonalization, personal accomplishment (MBI; Maslach and Jackson 1981)	5 weeks	x				x			4	2	0
Dicke et al. (2015)	Primary school, vocational track, intermediate track, and academic track	Germany	Emotional exhaustion (MBI; Maslach et al. 1996)	2.5 days			x			x		2	1	3

Table 1 (continued)

Study	School level	Country	Outcome	Intervention length	Approach							Risk of bias		
					Cognitive behavioral	Mindfulness/ meditation	Professional development	Psycho educational	Social support	Socio-emotional skills	Other	Low risk	Unclear	High
Domitrovich et al. (2016)	Elementary school	USA	Emotional exhaustion, personal accomplishment (MBI; Maslach et al. 1996)	5 days						x		6	0	0
Dreyer (2012)	College (higher education)	New Zealand	Emotional exhaustion (PBQ; Pines et al. 1981)	10 weeks							Physical exercise	4	2	0
Ebert et al. (2014)	Mixed sample	Germany	Emotional exhaustion, depersonalization, personal accomplishment (MBI-D; Büssing and Perrar 1992)	5 weeks	x							5	1	0
Emery (2011)	Elementary and secondary school	USA	Burnout (MBI-ES; Maslach et al. 1996)	1 day				x		x		2	1	3
Flook et al. (2013)	Elementary school	USA	Emotional exhaustion, depersonalization, personal accomplishment (MBI-ES; Maslach et al. 1996)	8 weeks	x	x						3	3	0
Frank et al. (2013)	High school	USA	Emotional exhaustion, depersonalization, personal accomplishment (MBI-ES; Maslach et al. 1996)	8 weeks		x						3	1	2
Harris et al. (2016)	Middle school	USA	Emotional exhaustion, depersonalization, personal accomplishment (MBI-ES; Maslach and Jackson 1981)	16 weeks		x						4	2	0
Jennings et al. (2013)	Regular education, special education, and art education	USA	Emotional exhaustion, depersonalization, personal accomplishment (MBI-ES; Maslach et al. 1996)	5 days		x			x	x		5	1	0
Johnson and Naidoo (2013)	Primary school	South Africa	Burnout (CBI; Kristensen and Borritz 1999)	–		x						1	2	3

Table 1 (continued)

Study	School level	Country	Outcome	Intervention length	Approach							Risk of bias		
					Cognitive behavioral	Mindfulness/ meditation	Professional development	Psycho educational	Social support	Socio-emotional skills	Other	Low risk	Unclear	High
Porter (1999)	Middle school, junior high, high school, or alternative school levels	USA	Emotional exhaustion, depersonalization, personal accomplishment (MBI-ES; Iwanicki and Schwab 1981; Gold 1984)	5 sessions (unclear time correspondence)	x			x				1	3	2
Roeser et al. (2013)	Mixed sample (approx. 50% elementary school)	Canada and USA	Burnout (MBI; Maslach et al. 2001)	8 weeks		x						4	2	0
Siu et al. (2014)	Primary and secondary school	China	Emotional exhaustion (MBI-HSS; Maslach et al. 1996)	2.5 days							Positive psychology	3	1	2
Unterbrink et al. (2012)	Grammar schools and secondary modern schools	Germany	Emotional exhaustion, depersonalization, personal accomplishment (MBI; Maslach et al. 1996)	10 months				x	x			5	1	0
Wolf et al. (2015)	Elementary school	Democratic Republic of the Congo	Burnout (MBI; Maslach et al. 1996)	1 school year			x					6	0	0

Risk of bias (study quality) was assessed using the six criteria (i.e., sequence generation, allocation concealment, blinding of outcome assessor, incomplete outcome data, selective outcome reporting, other potential threats to validity) of the Cochrane Collaboration tool. The three scores represent the total number of criteria on which each study presented low, unclear, or high risk of bias
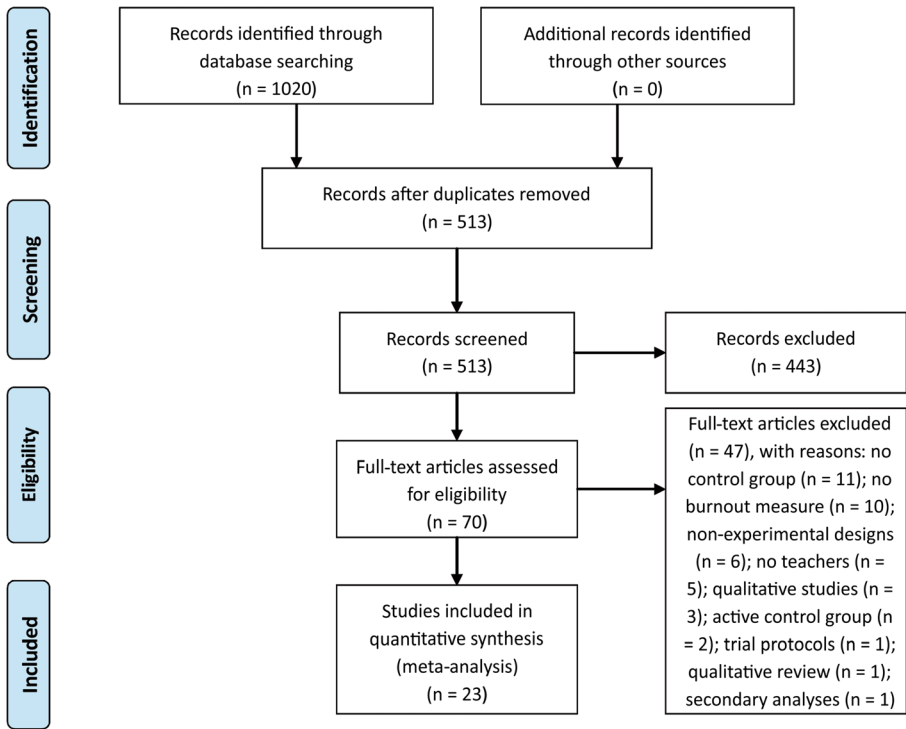
Fig. 1 Flow chart of the selection of studies, following the PRISMA statement

the software automatically implemented the algorithms. When we conducted our overall analyses, we combined all effect sizes coming from a single research study (all measures and all measurement moments). In our moderator analyses, we selected only the effect sizes relevant for that particular category (e.g., effect sizes for the exhaustion scale), and we combined all measurement moments reported in that particular study.
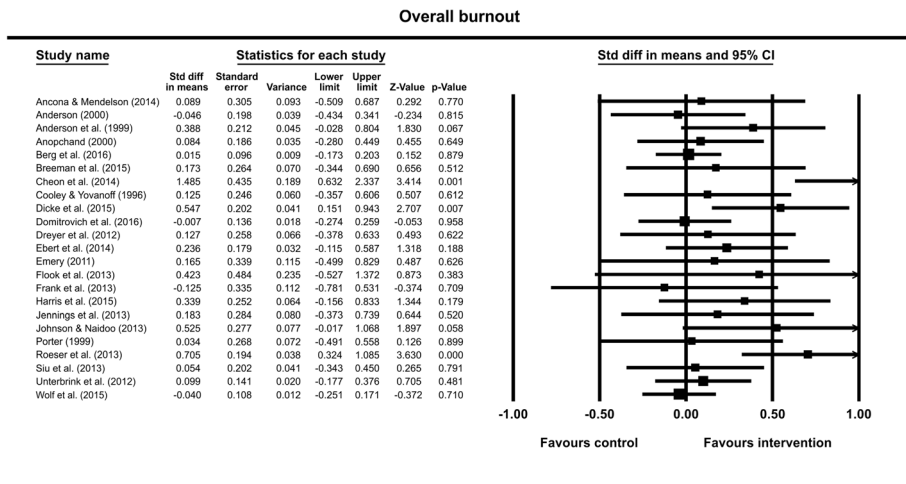

Fig. 2 Standardized effect sizes and forest plot for the entire sample of studies for overall burnout symptoms

Table 2 Overall effects of the interventions

Outcome	k	d	SE	Min d	Max d	Z	Q	I^2
Overall burnout	23	0.18	0.05	0.07	0.29	3.26**	34.42*	36.08
Emotional exhaustion	19	0.18	0.06	0.06	0.30	2.97**	27.15	33.69
Depersonalization	11	0.03	0.06	−0.08	0.14	0.53	4.86	0.00
Personal accomplishment	13	0.14	0.06	0.03	0.25	2.46**	13.02	7.83

k the number of studies included in the analysis, d the average effect size, SE standard error of the average effect size, min/max d the minimum and maximum limits of the confidence interval, Z the statistical test used for computing the significance of the average effect size, Q the statistical test used for the estimation of heterogeneity, I^2 the proportion of effect size variance that can be attributed to moderator variables

*Effect is statistically significant at $p < 0.05$; **effect is statistically significant at $p < 0.01$

Selection and Inclusion of Studies

As displayed in Fig. 1, the systematic search in electronic databases identified an initial pool of 1020 records out of which 513 were unique materials (pool of records after the exclusion of duplicates). After screening the abstracts of the initial set of materials, we identified 70 manuscripts as being potentially eligible and retrieved them in full text for further verification. In the next selection phase, we excluded 40 records (11 lacked a control group, 10 did not measure burnout or any of its components, 6 were non-experimental studies—e.g., cross-sectional—5 targeted other populations than teachers or included mixed professions, 3 were qualitative studies, 2 had an active control group, 1 was a trial protocol, 1 was an intervention framework proposal, and 1 article reported secondary analyses of a previously published study). From the remaining 30 records, 7 lacked (some of) the needed statistical estimates in order to compute the effect size or reported measuring burnout but without including the outcome in the analyses. For these cases, we contacted the corresponding authors and requested the needed information. We received no replies. Hence, the present systematic review summarizes the results from a sample of 23 materials (19 journal articles and 4 dissertations). We also conducted additional searches in the existing systematic and qualitative reviews, but we did not find any supplemental records.

Results

Characteristics of Included Studies

Almost half of the eligible studies ($n = 14$) included mixed samples of teachers, with the widest of them ranging from primary school to high school (Ancona and Mendelson 2014; Anderson 2000; Anderson et al. 1999; Cheon et al. 2014; Cooley and Yovanoff 1996; Dicke et al. 2015; Ebert et al. 2014; Emery 2011; Jennings et al. 2013; Harris et al. 2016; Porter 1999; Roeser et al. 2013; Siu et al. 2014; Unterbrink et al. 2012). Of the remaining studies, six were implemented on elementary school teachers (Anopchand 2000; Berg, et al. 2016; Domitrovich et al. 2016; Flook et al. 2013; Johnson and Naidoo 2013; Wolf et al. 2015), two were designed for middle school (Anderson 2000; Harris et al. 2016), one was conducted for high school teachers (Frank et al. 2013); another one included staff from a higher education institution (Dreyer 2012). There was also a study

Table 3 Intervention effectiveness for different intervention approaches

Category	K	d	SE	Min d	Max d	Z	Q	I^2
Emotional exhaustion								
Cognitive behavioral	5	0.20	0.11	−0.00	0.41	1.94*	2.40	0.00
Mindfulness/meditation	6	0.31	0.12	0.08	0.54	2.64**	4.23	0.00
Social-emotional approach	3	0.04	0.11	−0.18	0.26	0.40	1.52	0.00
Psychoeducational approach	3	0.09	0.11	−0.12	0.29	0.83	0.42	0.00
Social support	3	0.12	0.11	−0.10	0.34	1.03	0.12	0.00
Professional development	4	0.46	0.29	−0.11	1.03	1.59	16.63**	81.96
Depersonalization								
Cognitive behavioral	5	−0.01	0.10	−0.22	0.19	−0.12	0.62	0.00
Mindfulness/meditation	5	0.17	0.13	−0.08	0.41	1.35	2.49	0.00
Social-emotional approach	–	–	–	–	–	–	–	–
Psychoeducational approach	3	−0.04	0.11	−0.24	0.17	−0.33	0.27	0.00
Social support	3	−0.04	0.11	−0.26	0.19	−0.31	0.69	0.00
Professional development	–	–	–	–	–	–	–	–
Personal accomplishment								
Cognitive behavioral	5	0.17	0.14	−0.10	0.45	1.23	6.50	38.47
Mindfulness/meditation	5	0.28	0.14	−0.00	0.56	1.96*	4.90	18.32
Social-emotional approach	2	0.08	0.12	−0.17	0.32	0.61	1.00	0.29
Psychoeducational approach	3	0.07	0.13	−0.19	0.33	0.52	2.87	30.41
Social support	3	0.27	0.11	0.05	0.49	2.43*	0.16	0.00
Professional development	2	0.10	0.13	−0.16	0.36	0.74	1.34	25.19

k the number of studies included in the analysis, d the average effect size, *SE* standard error of the average effect size, *min/max d* the minimum and maximum limits of the confidence interval, *Z* the statistical test used for computing the significance of the average effect size, *Q* the statistical test used for the estimation of heterogeneity, I^2 the proportion of effect size variance that can be attributed to moderator variables

*Effect is statistically significant at $p < 0.05$; **effect is statistically significant at $p < 0.01$

in which participated teachers were working with children with psychiatric disorders, namely special primary school children (Breeman et al. 2016).

Table 4 Intervention effectiveness for teaching level

Category	k	d	SE	Min d	Max d	Z	Q	I^2
Emotional exhaustion								
Primary	5	0.01	0.07	−0.13	0.14	0.07	0.54	0.00
Middle	2	0.10	0.16	−0.21	0.40	0.62	0.81	0.00
Mixed	10	0.32	0.10	0.12	0.51	3.19**	16.98*	47.00
Depersonalization								
Primary	2	0.02	0.09	−0.16	0.21	0.23	0.00	0.00
Middle	2	0.17	0.16	−0.14	0.48	1.06	1.04	3.91
Mixed	6	0.02	0.08	−0.15	0.18	0.19	2.22	0.00
Personal accomplishment								
Primary	4	0.12	0.11	−0.11	0.34	1.04	4.85	38.10
Middle	2	0.08	0.27	−0.45	0.61	0.29	2.85	64.85
Mixed	6	0.25	0.08	0.09	0.41	3.01**	0.86	0.00

k the number of studies included in the analysis, d the average effect size, *SE* standard error of the average effect size, *min/max d* the minimum and maximum limits of the confidence interval, *Z* the statistical test used for computing the significance of the average effect size, *Q* the statistical test used for the estimation of heterogeneity, I^2 the proportion of effect size variance that can be attributed to moderator variables

*Effect is statistically significant at $p < 0.05$; **effect is statistically significant at $p < 0.01$

Table 5 Intervention effectiveness for time lag

Category	k	d	SE	Min d	Max d	Z	Q	I^2
Emotional exhaustion								
<1 month	13	0.04	0.05	−0.06	0.15	0.83	1.95	0.00
1–3 months	5	0.46	0.17	0.14	0.79	2.77**	10.41*	61.57
>3 months	3	0.68	0.24	0.21	1.15	2.82**	5.66	64.68
Depersonalization								
<1 month	8	−0.01	0.06	−0.14	0.11	−0.21	1.67	0.00
1–3 months	3	0.16	0.12	−0.08	0.39	1.33	1.58	0.00
>3 months	–	–	–	–	–	–	–	–
Personal accomplishment								
<1 month	10	0.10	0.07	−0.03	0.23	1.54	10.65	15.45
1–3 months	3	0.29	0.12	0.05	0.52	2.39*	0.15	0.00
>3 months	–	–	–	–	–	–	–	–

k the number of studies included in the analysis, d the average effect size, SE standard error of the average effect size, min/max d the minimum and maximum limits of the confidence interval, Z the statistical test used for computing the significance of the average effect size, Q the statistical test used for the estimation of heterogeneity, I^2 the proportion of effect size variance that can be attributed to moderator variables

*Effect is statistically significant at $p < 0.05$; **effect is statistically significant at $p < 0.01$

As can be seen from Table 1, based on the reported descriptions for the implemented interventions, we identified the following six broader approaches: (1) cognitive behavioral ($n = 5$), (2) mindfulness/meditation ($n = 9$), (3) professional development ($n = 5$), (4) psychoeducational ($n = 4$), (5) social support ($n = 3$), and (6) socio-emotional skills ($n = 4$). In the first category, we included studies that explicitly mentioned applying a *cognitive behavioral* framework (e.g., rational emotive behavior therapy; Anderson 2000) or interventions that implied cognitive/thoughts' restructuring, goal setting, and planning and other similar techniques (e.g., problem-solving training—Ebert et al. 2014). The *mindfulness/meditation* approach comprised all the interventions that had a mindfulness training or/and any type of meditation (e.g., mantra meditation and pranayama—Anderson et al. 1999). We labeled as *professional development* any program that was aimed at training the teachers with skills for student interaction and classroom management (e.g., Good Behavior Game— Breeman et al. 2016; training in instructional practices—Wolf et al. 2015). In *psychoeducational*, we included all the interventions that comprised lectures about burnout, stress, or mental health in general (e.g., an overview of the state of the art on the prevention of burnout in teachers—Porter 1999; information concerning stress biology—Unterbrink et al. 2012). Examples of *social support* approaches implied organizing participants in peer collaboration programs (e.g., Cooley and Yovanoff 1996) or using group work in the intervention (e.g., Unterbrink et al. 2012). In the *socio-emotional skills* category, we grouped the studies that were less explicit regarding their interventions' framework but stated that it was intended to improve such skills in teachers (e.g., Jennings et al. 2013). Finally, there were three studies with unique approaches (i.e., the *other* category). More precisely, Anopchand (2000) used the cathartic approach of expressive writing; Dreyer (2012) used a physical exercise program, and Siu et al. (2014) had a positive psychology approach.

Moreover, the tested interventions were also in various lengths, starting from less than 1 month ($n = 8$), between 1 and 3 months ($n = 6$), from 3 to 5 months ($n = 2$), and going up to 1 year (9–12 months; $n = 4$). In three cases, the intervention's length was not reported or unclear (Anderson 2000; Johnson and Naidoo 2013; Porter 1999). Last

Table 6 Intervention effectiveness for length of the intervention

Category	k	d	SE	Min d	Max d	Z	Q	I^2
Emotional exhaustion								
<1 month	7	0.17	0.08	0.02	0.32	2.15*	6.57	8.72
1–3 months	5	0.33	0.12	0.09	0.57	2.65**	3.62	0.00
3–5 months	2	0.83	0.60	−0.36	2.01	1.37	5.81*	82.78
8–12 months	3	0.03	0.08	−0.12	0.18	0.36	0.50	0.00
Depersonalization								
<1 month	2	0.08	0.15	−0.21	0.38	0.55	0.12	0.00
1–3 months	4	0.01	0.14	−0.26	0.28	0.09	2.12	0.00
3–5 months	–	–	–	–	–	–	–	–
8–12 months	2	−0.00	0.08	−0.16	0.15	−0.03	0.18	0.00
Personal accomplishment								
<1 month	3	0.13	0.10	−0.06	0.33	1.32	1.74	0.00
1–3 months	4	0.26	0.19	−0.11	0.62	1.39	4.74	36.72
3–5 months	–	–	–	–	–	–	–	–
8–12 months	3	0.13	0.09	−0.04	0.30	1.49	2.39	16.31

k the number of studies included in the analysis, d the average effect size, SE standard error of the average effect size, min/max d the minimum and maximum limits of the confidence interval, Z the statistical test used for computing the significance of the average effect size, Q the statistical test used for the estimation of heterogeneity, I^2 the proportion of effect size variance that can be attributed to moderator variables

*Effect is statistically significant at $p < 0.05$; **effect is statistically significant at $p < 0.01$

but not least, it is worth noticing that slightly more than half of the included studies were conducted in the USA ($n = 14$; Roeser et al. (2013) had a mixed America-Canadian sample). Next, in a descending order, there were studies conducted in Western Europe ($n = 4$; 3 in Germany and 1 in the Netherlands), Asia ($n = 2$; 1 in China and 1 in South Korea), Africa ($n = 2$; 1 in the Democratic Republic of Congo and 1 in South Africa), and finally, one study was done in New Zeeland. As it can be observed, the large majority of the studies were conducted in the western (especially North-American) culture. This aspect is important since it represents a major bias for the cross-cultural (external) validity of the results.

Overall Effects on Burnout Symptoms

From the 23 studies included in the quantitative analyses, 4 measured only overall burnout symptoms (Emery 2011; Johnson and Naidoo 2013; Roeser et al. 2013; Wolf et al. 2015). All the remaining 19 studies covered emotional exhaustion, 11 of these also measured depersonalization and 13 captured personal accomplishment (see Table 1 and Fig. 2 for details).

Both the overall mean ES and also the ones for each burnout component were weak in magnitude (Table 2). The strongest mean effect size was the overall one ($d = 0.18$; 95% CI 0.07, 0.29; $p = 0.001$) and the one for emotional exhaustion ($d = 0.18$; 95% CI 0.06, 0.30; $p = 0.003$). Not far from the aforementioned values was the mean effect size for personal accomplishment ($d = 0.14$; 95% CI 0.03, 0.25; $p = 0.014$), while the weakest effect was reached for depersonalization ($d = 0.03$; 95% CI −0.08, 0.14; $p = 0.599$), which was almost null.

We assessed the heterogeneity of the results using the Q statistic, the I^2 index, and also based on the width of the confidence interval (CI). The I^2 index showed some between-study variance for the overall burnout (36.1%) and emotional exhaustion (33.7%), while for
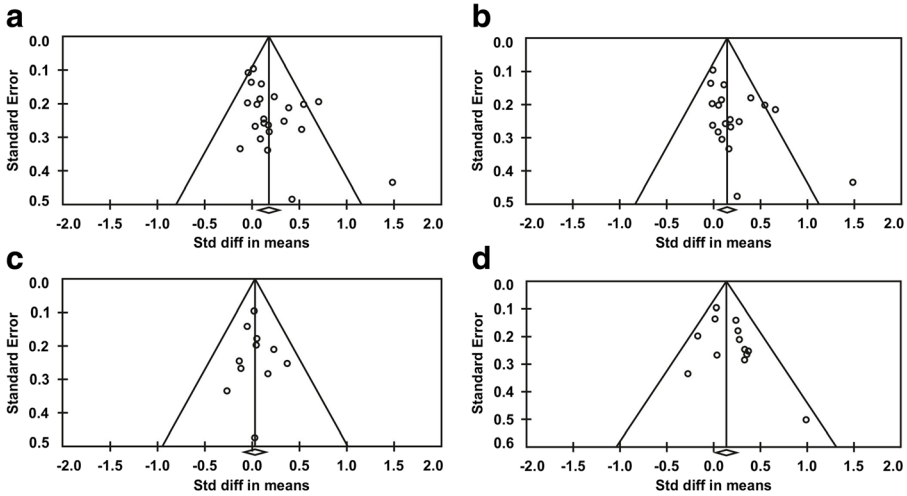
Fig. 3 Funnel plots for publication bias. **a** Overall burnout. **b** Emotional exhaustion. **c** Depersonalization. **d** Personal accomplishment

depersonalization and personal accomplishment, the included effects were quite homogeneous (0.0 and 7.8%). Moreover, for the overall burnout effect, the Q test was statistically significant, thus converging with the I^2, and corroborated with a CI that stretched from very small to small effect sizes; we can conclude that there was some degree of heterogeneity. The situation is similar also for emotional exhaustion (except that Q is non-significant).

Intervention Approach as Moderator

As can be seen in Table 3, emotional exhaustion was significantly alleviated by the cognitive behavioral approaches ($d = 0.20$; 95% CI −0.00, 0.41; $p = <0.05$) and those comprising mindfulness/meditation techniques ($d = 0.31$; 95% CI 0.08, 0.54; $p = <0.01$). All the remaining four approaches produced non-significant effects. Moreover, except for the studies based on professional development ($Q = 16.63$, $p < 0.01$; $I^2 = 81.9\%$), all the other ones had very homogeneous effects ($I^2 = 0.0\%$ in all cases).

None of the identified approaches seem to have worked in alleviating depersonalization symptoms. All the effects were very small in magnitude and non-significant. Moreover, the results' heterogeneity is also very low ($I^2 = 0.0\%$ in all cases).
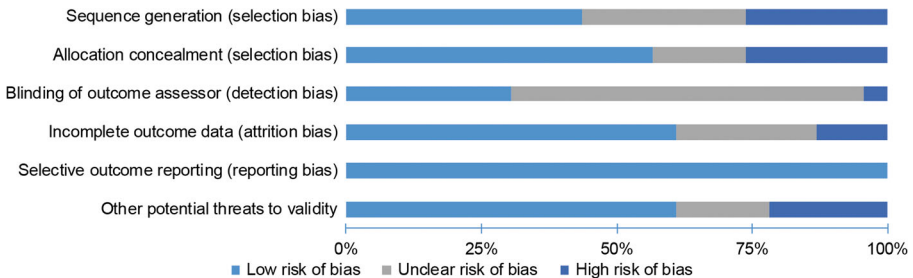

Fig. 4 Risk of internal bias summary

Personal accomplishment was significantly increased by mindfulness/meditation ($d = 0.28$; 95% CI −0.00, 0.56; $p < 0.05$) and also by social support ($d = 0.27$; 95% CI 0.05, 0.49; $p < 0.05$). For all the other approaches, the effects were non-significant. Moreover, as indexed by Q and I^2 statistics, the results have low heterogeneity.

However, moving beyond the statistical significance of the results, we consider worth mentioning that the CIs for all the effects (regardless the outcome or the approach) are overlapping. Hence, there is practically no difference in effectiveness between the identified approaches.

Teaching Level as Moderator

After grouping the studies based on their target populations, there are only two statistically significant effects. The interventions that included mixed samples of teachers were significantly effective in alleviating participants' emotional exhaustion ($d = 0.32$, 95% CI 0.12, 0.51; $p < 0.01$) and in increasing their personal accomplishment ($d = 0.25$, 95% CI 0.09, 0.41; $p < 0.01$). However, as can be seen from Table 4, the 95% confidence intervals of the effect sizes for all the categories of teachers are largely overlapping. Hence, we cannot conclude that there is actually a significant difference in interventions' effectiveness based on the sample composition.

Time Lag for the Outcome Measures as Moderator

Previous reviews suggested that intervention effects can vary in intensity at different time points (Awa et al. 2010). We also tested this idea by conducting separate analyses on burnout and the components of burnout that were assessed at post-test and follow-up. Follow-up means more than 1 month or more than 3 months after the intervention.

Results presented in Table 5 show that the average effect sizes have different values throughout the time. In the case of all burnout components, the post-intervention effects (i.e., in the first month following the intervention) are almost null, statistically non-significant, and highly homogeneous. Interestingly, medium-sized and significant effects were reported by the few research papers that measured the effectiveness at more than 1 month after the end of the intervention. More precisely, emotional exhaustion significantly decreased in the studies that measured it between 1 and 3 months after the intervention ($d = 0.46$; 96% CI 0.14, 0.79; $p < 0.01$) and also at more than 3 months ($d = 0.68$; 96% CI 0.21, 1.15; $p < 0.01$). Personal accomplishment was also successfully manipulated in the studies that measured it between 1 and 3 months from the intervention ($d = 0.29$; 95% CI 0.05, 0.52; $p < 0.05$).

Intervention Duration as Moderator

Intervention duration is another potential moderator that we investigated (Table 6). Our results suggested that interventions lasting between 1 and 3 months reported homogeneous and significant results on emotional exhaustion ($d = 0.33$, 95% CI 0.09, 0.57; $p < 0.01$). Similar to the results reported by Maricuțoiu et al. (2016), interventions shorter than a month reported smaller effect sizes ($d = 0.17$ for exhaustion, $d = 0.13$ for personal accomplishment), as

compared with interventions that lasted between 1 and 3 months ($d = 0.33$ for exhaustion; $d = 0.26$ for personal accomplishment). In the case of depersonalization, interventions with different lengths did not have different effectiveness.

Publication Bias

For the overall effects ($k = 23$), there were small traces of publication bias. Egger's test was statistically significant (intercept = 1.59; 95% CI 0.37, 2.82; $p = 0.013$), but the Duval and Tweedie trim and fill procedure did not impute any studies. Moreover, the funnel plot has a symmetric appearance except for the absence of any higher negative ES (Fig. 3a). For emotional exhaustion ($k = 19$), the Egger's test was also significant (intercept = 1.54; 95% CI 0.12, 2.95; $p = 0.036$), but the Duval and Tweedie trim and fill procedure signaled not bias. The visual inspection of the funnel plot also reveals a generally symmetric distribution, except for the absence of higher negative effect sizes (Fig. 3b). Depersonalization ($k = 11$) was the least biased category of effects. Egger's test was not significant (intercept = 0.09; 95% CI −1.13, 1.32; $p = 0.866$); the Duval and Tweedie trim and fill procedure did not impute any studies, and the funnel plot was clearly symmetrical (Fig. 3c). Personal accomplishment was also symmetrically represented (Fig. 3d). The absence of the threat of publication bias was congruently supported by the non-significant Egger's test (intercept = 1.15; 95% CI −0.33, 2.63; $p = 0.114$) and the Duval and Tweedie trim and fill procedure (0 trimmed studies).

Quality of the Included Studies

Sequence generation was at low risk of bias for 10 out of the 23 included studies; for 7 studies, the risk was unclear, while the remaining 6 had a high risk (Fig. 4). Allocation concealment was at low risk for more than half of the studies ($n = 13$); four had insufficient information for this criterion, and six were at high risk. Masking of participants was not possible since all studies had a passive control group (e.g., waiting list); hence, we assessed only whether the outcome assessors were masked. Our assessment revealed that this was possible for only 7 studies, while for a majority of 15, it was unclear if the blinding of assessors was done. Moreover, one study was clearly at high risk. Attrition bias (i.e., incomplete outcome data) was at low risk in 14 studies, unclear in 6, and at high risk in 3. As regarding the reporting bias (i.e., selective outcome reporting), all 23 studies were considered to be at low risk.

For the majority of the studies ($n = 14$), we could not detect other sources of bias; in four cases, it was unclear if there are other potential threats, and five cases were identified at high risk of additional biases. More precisely, Ancona and Mendelson (2014) did not account for the nesting structure of their data (i.e., teachers within schools). Berg et al.'s (2016) study could be biased regarding the statistical power since the sample was small and became smaller after group allocation. Breeman et al. 2016) had teachers who refused study enrollment because of experiencing burnout; hence, this could signal a selection bias since a group of participants with high relevance regarding the outcome of interest (i.e., burnout) was lost. Emery (2011) reports potential problems regarding the operationalization of one of the outcomes (i.e., helping values), and even if this is not related to burnout, we still consider important to take into account any possible threat to the operationalizations' validity. Finally, Johnson and Naidoo (2013) reported between-group variations regarding participants' gender distribution, an imbalance which represents a possible selection bias.

Overall, study quality was optimal. Slightly more than two thirds of the studies ($n = 16$) were at low risk of bias for at least three or more criteria (50% or more of the criteria), out of which three met all six quality criteria (see Table 1). We also conducted a meta-regression with study quality (total number of criteria with low risk) as a predictor for the overall effectiveness. The effect was non-significant ($b = -0.028$; SE = 0.025; 95% CI −0.08, 0.02; $p = 0.257$). Hence, study quality had a negligible impact on the effectiveness of the interventions in alleviating burnout symptoms.

Discussion

Today, researchers invest effort in developing and testing interventions aimed at reducing or preventing teacher burnout (Maslach et al. 2001). Although previous reviews (Maricuțoiu et al. 2016) did not explore whether interventions work differently from one occupation to another, the idea of conducting analyses on separate professions is gaining the researchers' interest (Panagioti et al. 2016). Teachers are different from other occupational categories because they have teaching-specific demands and resources (McCarthy et al. 2016); therefore, the primary objective of this study was to review the existing evidence regarding the effectiveness of interventions on teacher burnout. To achieve this goal, we carried out a systematic literature search from which we analyzed 23 controlled studies, most of them conducted in the USA.

Our overall results indicated that intervention effectiveness is generally small and similar to previous reviews (Maricuțoiu et al. 2016). The existing approaches are effective for emotional exhaustion and personal accomplishment. With regard to depersonalization, the effectiveness of these interventions is almost null. These effects are highly homogeneous in the case of depersonalization and personal accomplishment, while moderate levels of heterogeneity were present in the case of overall burnout and emotional exhaustion. One possible explanation for this difference is that Maricuțoiu et al. (2016) integrated all types of occupational categories, while the present review focused on a specific professional group.

However, given that about 30% of the between-study variance can be attributed to moderator variables, we investigated potential moderators of intervention efficacy. Similar to previous reviews (Richardson and Rothstein 2008; Maricuțoiu et al. 2016), we grouped interventions based on their approach, and we conducted separate analyses on the three burnout components. Interestingly, interventions based on mindfulness reported significant and homogeneous effects on exhaustion and personal accomplishment. In addition, mindfulness-based interventions also reported a small effect on depersonalization, but it did not reach statistical significance due to the small number of studies. Other approaches that had a significant impact were cognitive behavioral interventions (significant effects in the case of exhaustion) and interventions based on social support (significant effects in the case of personal accomplishment). All these results are of particular importance because previous reviews did not report statistically and homogeneous results regarding the effectiveness of a particular intervention approach. The existing literature does not allow for the investigation of the change mechanisms that explain the effectiveness of these interventions in reducing teacher exhaustion. Therefore, more studies are needed in order to achieve an improved understanding of these change mechanisms.

Another interesting finding of our analyses on groups of intervention approaches is that some approaches (i.e., psychoeducational interventions, social-emotional interventions) had almost null effects on all burnout components. Regarding interventions based on the enhancing the social-emotional skills, it is possible that their effects can be more distant in time. For example, Spilt et al. (2011) suggested that high-quality teacher-student relations can generate teacher well-being in the long run. From this perspective, interventions based on social-emotional skills might have improved the teacher-student relations, but the effects on teacher well-being are simply not observable in the post-intervention assessment. Therefore, follow-up measures are needed in order to conclude regarding the effectiveness of social-emotional interventions.

Finally, the professional development approach needs further investigation because it seems to have a large effect on exhaustion and because further research is needed to clarify whether it has an effect on depersonalization or on personal accomplishment. The professional development approach aims to enhance students' communication and interpersonal skills (Berg et al. 2016; Breeman et al. 2016), which, in turn, can reduce teachers' burnout. Teacher burnout is a secondary outcome of this type of intervention: following the intervention, students change their behavior (i.e., primary outcome), which leads to improvements in teacher burnout levels (i.e., secondary outcome). Therefore, we encourage future research to investigate whether teacher burnout will evolve in time, although post-intervention changes are not statistically significant.

In additional moderator analyses, we grouped the research studies based on the participants' teaching level, based on the time lag between the end of the intervention and the moment of effectiveness assessment, and based on the overall length of interventions. Regarding teaching level, most studies used mixed samples (i.e., teachers from different levels) and reported larger effect sizes, as compared with interventions on primary school teachers or interventions on middle school teachers. These differences were present in the case of exhaustion and personal accomplishment, but not in the case of depersonalization. Similar to the conclusions of Maricuțoiu et al. (2016), interventions that lasted between 1 and 3 months reported stronger effect sizes on exhaustion and personal accomplishment, as compared with interventions with different lengths.

Finally, our analyses regarding the time lag between the end of intervention and the assessment moment provided unclear results. On the one hand, it seems that effect sizes are close to null when the time lag is less than a month (i.e., at post-intervention assessment). On the other hand, the few research studies that measured intervention effectiveness using time lags larger than 1 month reported large effect sizes. However, these latter studies did not measure burnout immediately after the intervention. Therefore, it is unclear whether these studies had large or small effect sizes immediately after the end of the intervention.

The quality (i.e., the risk of bias) of the included studies can be generally described as optimal. More than two thirds of the sample of studies had low risk for at least 50% of the quality criteria. Adherence to highly standardized study designing and reporting frameworks (e.g., CONSORT; Boutron et al. 2008) should be strongly encouraged in order to further minimize the risk of bias for future studies. Also, it is important to note that study quality seemed to be unrelated to the effectiveness of the intervention. This finding certifies the validity of our results.

Limitations

We should mention some limitations of this meta-analysis. Firstly, an important limitation is the level at which an intervention is conducted (individual level vs. organizational level). All

studies included in this review are individual-level interventions. Some authors (e.g., Maslach et al. 2001) suggested that by focusing excessively on individual behavior change will not produce persistent improvements in all psychological outcomes. Therefore, future research should investigate the effectiveness of organizational interventions (e.g., changes in the organizational politics) in educational settings.

Moreover, it is unclear whether the participants actually experienced burnout and needed the intervention in the first place. In most cases, participants enrolled in the programs voluntarily, not because they were diagnosed as suffering from the burnout syndrome. As a consequence, the post-intervention results were not very different from the pre-intervention evaluations, and this could explain why the effect sizes identified here are generally small.

In some moderation analyses, a possible limitation is represented by the fact that subgroup analyses are mostly based on small samples of studies. Because of the low statistical power, subgroup analyses (e.g., the intervention-type categories) should be treated with reservations.

Although we did not use any language restrictions, another possible limitation of this review is that all included papers were written in English. This is due to the fact that about half of the studies (14 out of 23 studies) were conducted in the USA, and the remaining studies described interventions conducted in Europe (four papers), Asia (three papers), and Africa (two papers). Therefore, more research is needed to ensure higher levels of generalization.

Finally, we had seven research papers with incomplete statistical estimates needed to compute the effect size. Although we contacted the authors by e-mail, we excluded these studies because we did not receive a response. Although this is an important limitation, we believe that it is unlikely that these studies could have made a considerable impact on the overall results.

Conclusions and Implications for Future Research

Despite the overall significant effect of these approaches, the small effect size raises a series of problems that should be addressed by future research studies. Firstly, most researchers did not use interventions tailored for the educational environment. For example, the teacher-specific stressors can differ from one educational level to another (i.e., primary school, middle school, or high school). As a consequence, studies that used only primary school teachers reported almost null results. Secondly, our small effect sizes suggest that teacher burnout has causes that are not addressed by these interventions. Most interventions were developed using a general model of stress and did not address teacher-specific stressors. For example, psychoeducational or supportive collegial interactions are not helpful when teachers are emotionally drained. Therefore, researchers should focus more on stressors specific to the teaching environment, and these stressors should be addressed by future interventions.

Future interventions can be tailored using teacher-specific stress models (e.g., the classroom appraisal of resources and demands model—McCarthy et al. 2016), and these approaches should provide improved results regarding teacher burnout. In addition, future interventions can address change mechanisms specific to the teaching environment. For example, although there is substantial evidence that classroom management self-efficacy is related to teacher burnout (Aloe et al. 2014), the interventions included in this review did not address this particular self-regulation mechanism. Therefore, future interventions should investigate whether the enhancement of classroom management self-efficacy can improve teacher well-being (in general) and decrease teacher burnout (in particular).

Besides burnout, these interventions also aimed at improving other work-related variables. Specifically, significant improvements were attained with regard to anxiety symptoms (Roeser et al. 2013; Johnson and Naidoo 2013), depressive symptoms (Roeser et al. 2013; Ebert et al. 2014), mindfulness (Roeser et al. 2013; Flook et al. 2013; Jennings et al. 2013), self-efficacy (Jennings et al. 2013; Ebert et al. 2014; Domitrovich et al. 2016), and job satisfaction (Cooley and Yovanoff 1996; Siu et al. 2014; Wolf et al. 2015). Although these interventions did not significantly diminish burnout, they managed in some way to increase other well-being variables. Therefore, future studies should also assess their efficacy using a larger panel of well-being variables.

References

References marked with an * were included in the review

Aloe, A. M., Amo, L. C., & Shanahan, M. E. (2014). Classroom management self-efficacy and burnout: a multivariate meta-analysis. *Educational Psychology Review, 26*, 101–126. doi:10.1007/s10648-013-9244-0.

*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*, 156–170. doi: 10.1080/1754730X.2014.920135.

*Anderson, S. L. (2000). *Effects of a rational-emotive behavior therapy intervention on irrational beliefs and burnout among middle school teachers in the state of Iowa* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (UMI No 9978792).

*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*, 3–25.

*Anopchand, N. K. (2000). *The impact of expressive writing on prospective teachers' sense of efficacy, stress, burnout, and satisfaction with teaching* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (UMI No 9976516).

Awa, W. L., Plaumann, M., & Walter, U. (2010). Burnout prevention: a review of intervention programs. *Patient Education and Counseling, 78*, 184–190. doi:10.1016/j.pec.2009.04.008.

*Berg, J. K., Bradshaw, C. P., Jo, B., & Ialongo, N. S. (2016). Using complier average causal effect estimation to determine the impacts of the Good Behavior Game preventive intervention on teacher implementers. *Administration and Policy in Mental Health and Mental Health Services Research*, 1–14. doi: 10.1007/s10488-016-0738-1.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis (version 2) [computer software]*. Englewood: Biostat.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester: Wiley.

Boutron, I., Moher, D., Altman, D. G., Schulz, K. F., & Ravaud, F. (2008). Methods and processes of the CONSORT group: example of an extension for trials assessing nonpharmacologic treatments. *Annals of Internal Medicine, 148*, W–66. doi:10.7326/0003-4819-148-4-200802190-00008-w1.

*Breeman, L. D., van Lier, P. A., Wubbels, T., Verhulst, F. C., van der Ende, J., Maras, A., ... & 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*, 1–12. doi: 10.1177/1098300715593466.

Büssing, A., & Perrar, K. M. (1994). Burnout measurement. Study of a German version of the Maslach Burnout Inventory (MBI-D). Pflege Zeitschrift, 47, suppl-20.

*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*, 331–346. doi: 10.1123/jsep.2013-0231.

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

*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. doi: 10.1016/j.tate.2015.01.013.

*Domitrovich, C. E., Bradshaw, C. P., Berg, J. K., Pas, E. T., Becker, K. D., Musci, R., ... & 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*, 325–337. doi: 10.1007/s11121-015-0618-z.

*Dreyer, L. (2012). Effects of a 10-week high-intensity exercise intervention on college staff with psychological burnout and multiple risk factors. *The ICHPER-SD Journal of Research in Health, Physical Education, Recreation, Sport and Dance, 7*, 27–33.

*Ebert, D. D., Lehr, D., Boß, L., Riper, H., Cuijpers, P., Andersson, G., ... & 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, 40*, 582–596. doi: 10.5271/sjweh.3449.

*Emery, D. W. (2011). *Crisis in education: a call to ACT* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (UMI No3466610).

*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*, 182–195.

*Frank, J. L., Reibel, D., Broderick, P., Cantrell, T., & Metz, S. (2013). The effectiveness of mindfulness-based stress reduction on educator stress and well-being: results from a pilot study. *Mindfulness, 6*, 208–216. doi: 10.1007/s12671-013-0246-2.

Gold, Y. (1984). The factorial validity of the Maslach burnout inventory in a sample of California elementary and junior high school classroom teachers. *Educational and Psychological Measurement, 44*, 1009–1016. doi:10.1177/0013164484444024.

Gwet, K. L. (2012). *Handbook of inter-rater reliability (Third edition).* Gaithersburg: Advanced Analytics, LLC.

*Harris, A. R., Jennings, P. A., Katz, D. A., Abenavoli, R. M., & Greenberg, M. T. (2016). Promoting stress management and wellbeing in educators: feasibility and efficacy of a school-based yoga and mindfulness intervention. *Mindfulness, 7*, 143–154. doi: 10.1007/s12671-015-0451-2.

Higgins, J. P., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions (Vol. 4).* West Sussex: John Wiley & Sons.

Iwanicki, E. F., & Schwab, R. L. (1981). A cross validation study of the Maslach burnout inventory. *Educational and Psychological Measurement, 41*, 1167–1174. doi:10.1177/001316448104100425.

*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*, 374–390.

*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*, 59–70. doi: 10.1177/0081246312474406.

Kristensen, T. S., & Borritz, M. (1999). *Copenhagen burnout inventory.* Copenhagen: National Institute of Occupational Health.

Maricuțoiu, L. P., Sava, F. A., & Butta, O. (2016). The effectiveness of controlled interventions on employees' burnout: a meta-analysis. *Journal of Occupational and Organizational Psychology, 89*, 1–27. doi:10.1111/joop.12099.

Maslach, C., & Jackson, S. E. (1981). *Maslach burnout inventory.* Palo Alto: Consulting Psychologists Press.

Maslach, C., & Jackson, S. (1993). *Maslach burnout inventory manual* (2nd ed.). Palo Alto: Consulting Psychologists Press.

Maslach, C., Jackson, S., & Schwab, R. L. (1986). *Maslach burnout inventory- educators survey.* Palo Alto: Consulting Psychologists Press.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *Maslach burnout inventory. Manual* (3rd ed.). Palo Alto: Consulting Psychologists Press.

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology, 52*, 397–422. doi:10.1146/annurev.psych.52.1.397.

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*, 577–603. doi:10.1007/s10648-015-9322-6.

Panagioti, M., Panagopoulou, E., Bower, P., Lewith, G., Kontopantelis, E., Chew-Graham, C., ... & Esmail, A. (2016). Controlled interventions to reduce burnout in physicians: a systematic review and meta-analysis. *JAMA Internal Medicine.* doi: 10.1001/jamainternmed.2016.7674.

Pines, A. M., Aronson, E., & Kafry, D. (1981). *Burnout: From tedium to personal growth.* New York: Free Press.

*Porter, S. A. (1999). Teacher burnout: a quasiexperimental study of psychoeducational group interventions (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (UMI No 9938253).

Richardson, K. M., & Rothstein, H. R. (2008). Effects of occupational stress management intervention programs: a meta-analysis. *Journal of Occupational Health Psychology, 13*, 69–93. doi:10.1037/1076-8998.13.1.69.

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

*Roeser, R. W., Schonert-Reichl, K. A., Jha, A., Cullen, M., Wallace, L., Wilensky, R., ... & 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*, 787–804. doi: 10.1037/a0O32O93.

Schaufeli, W. B., & Van Dierendonck, D. (2000). *Manual of the burnout scale Utrecht (UBOS) [Handleiding van de UBOS]*. Lisse: Swets & Zeitlinger.

*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*, 69–84. doi: 10.1037/aa03329l.

Spilt, J. L., Koomen, H. M. Y., & Thijs, J. T. (2011). Teacher wellbeing: the importance of teacher–student relationships. *Educational Psychology Review, 23*, 457–477. doi:10.1007/s10648-011-9170-y.

Stebbings, J., Taylor, I. M., Spray, C. M., & Ntoumanis, N. (2012). Antecedents of perceived coach interpersonal Behaviors: The coaching environment and coach psychological well- and ill-being. *Journal of Sport and Exercise Psychology, 34*, 481–502. doi:10.1123/jsep.34.4.481.

*Unterbrink, T., Pfeifer, R., Krippeit, L., Zimmermann, L., Rose, U., Joos, A., ... & 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*, 667–674. doi: 10.1007/s00420-011-0712-x.

*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. doi: 10.1016/j.tate.2015.08.002.