



# Reducing conflicts in school environments using restorative practices: A systematic review

Constanze Weber\*, Leen Vereenooghe

Faculty of Psychology and Sports Science, Bielefeld University, P.O. Box 10 01 31, D-33501 Bielefeld, Germany



## ARTICLE INFO

### Keywords:

Restorative practice  
Restorative justice  
School  
Effectiveness  
Suspensions  
Relationships in school

## ABSTRACT

Restorative practices (RP) aim to repair the harm caused by conflicts and could thus offer an alternative to 'traditional' retributive interventions. In recent years, RP have been increasingly used with the aim of reducing conflicts in school settings. This systematic review examines the effectiveness of RP in school environments in reducing conflicts. We searched PsycINFO, PsycARTICLES, ASSIA, Social Science Citation Index, and ERIC, for articles which measure the outcome of RP in school settings that met the following inclusion criteria: (1) participants could be students or teachers from primary, middle or secondary schools, (2) study designs delivered quantitative outcome data, (3) peer-reviewed publication. Reference lists and citations of included full-text articles were checked for further studies. There were no limits regarding publication date, language or study design. Two reviewers independently screened titles and abstracts, and in a second step all identified full-text articles. Information on the authors, year of publication, country, study design, participants, outcomes and results of the study conducted were extracted. The database search yielded 600 studies, of which 17 articles met the inclusion criteria. Several studies demonstrated a decrease in student suspensions and behaviour referrals. Some studies suggest that the use of RP in schools might improve bullying and student-teacher relationships, although evidence remains limited. The studies included in this review were mostly correlational and non-experimental designs, which limits drawing conclusions regarding causality and generalisability. Research studies with robust study designs, such as randomised controlled trials, are urgently needed.

## 1. Introduction

Restorative practices (RP) are increasingly being adopted by schools as an alternative way of responding constructively to conflict or discipline and behaviour issues (Anfara, Evans & Lester, 2013; Green, Willging, Zamarin, Dehaiman & Ruiloba, 2019), and can be regarded as a means to establish a more inclusive and favourable school culture (Cremin, 2007; Hendry, 2009; McCluskey, Lloyd, Kane, Riddell, Stead & Weedon, 2008). In origin, RP are derived from 'restorative justice' within the criminal justice system, where the focus shifts from retribution and punishment of an offence to repairing the harm – to both people and relationships – caused by this offence (Zehr, 2002). The change from a judicial to a school setting followed in the 1990s in Australia (Wachtel, 2013), after which the term restorative justice is replaced by restorative practice, and refers to a process of mediation or reconciliation between victims and offenders, and by extension other members of the school community (Hopkins, 2004).

The absence of a unified and operationalised definition of RP (Daly, 2002; Reimer, 2011; Sellman, Cremin & McCluskey, 2014) and the different terms for the methods used in the context of RP present

researchers and practitioners with a barrier in determining best practices. In general, RP are considered an approach which provides a philosophy and a framework of proactive and reactive methods toward building and restoring relationships in schools and thereby reducing conflicts (Gregory, Clawson, Davis & Gerewitz, 2015; Hendry, 2009; Wachtel, 2013). Therefore, RP may prevent harm through community-building, as well as by responding to conflict in ways that would repair wounded relationships (González, 2012; Kline, 2016). In schools, these relationships concern those between teachers and students, between students themselves, between staff, and between staff and parents. Proactive and community-building approaches focus on building emotional intelligence and reducing the likelihood of student rule breaking, for example by means of conducting proactive circles. Responsive approaches focus on addressing wrongdoing, repairing harm, and restoring relationships after conflicts occurred and include, amongst others, restorative meetings and restorative conferences (Gregory et al., 2015; Hendry, 2009). Depending on the type and severity of the conflict, informal to formal methods of the RP continuum are used (Costello, Wachtel & Wachtel, 2009) (Table 1). Table 1 presents an overview of frequently used proactive and reactive restorative methods along the informal-formal continuum. Meanwhile, the SaferSanerSchools™ model includes

\* Corresponding author.

E-mail address: [constanze.weber@uni-bielefeld.de](mailto:constanze.weber@uni-bielefeld.de) (C. Weber).

**Table 1**  
Methods of restorative practices.

| Method                 | Description  |
|------------------------|--|
| Check-In/Check-Out     | A go-around in which each student responds to a question or a statement. Used at the beginning or the end of a class or day.   |
| Proactive Circles      | A routine (daily or weekly) classroom process run by the teacher. Generally, students sit in a circle and respond in a go-around technique to a question or statement that helps build community and create classroom norms.   |
| Affective Statements   | A statement containing a feeling, which is used in response to negative or positive events in the classroom or school.   |
| Restorative Questions  | Questions used in response to harm caused and intended to foster a process of reflection, e.g. 'Who has been affected by what you have done?' or 'What do you think you need to do to make things right?'  |
| Restorative Chat       | A dialogue between two persons addressing low-level (potentially) harmful behaviors. Lasts between 2 and 5 min and is held as soon as possible after harm was caused.  |
| Peer Mediation         | Two trained student mediators assisting fellow students, who are in conflict, to resolve their problems. The mediators work in pairs. Normally takes place in a private space, but close to the help of an adult.  |
| Restorative Meeting    | An interaction involving a teacher, the person responsible for causing harm and the person who has been harmed (or the two students who harmed each other). Restorative Meetings are used to address relatively minor incidents, if possible, within the same day. Likely duration is 15 to 60 min, including preparation. |
| Restorative Circle     | Used when three or more people have caused harm or have been harmed. Students sit in a circle and address who has been harmed and what needs to be done to make things right. The likely duration is usually 15 to 60 min, including preparation.  |
| Restorative Conference | Formal response to the most serious and complex incidents using a scripted approach to facilitate accountability and repair harm. All people involved and affected by the incident participate. Formal agreements are made. Likely duration (including any preparation) is around two hours.                               |

a few more elements, eleven in total ([International Institute for Restorative Practices, 2011](#)).

In a responsive restorative intervention, those affected by a conflict come together to identify how people were affected by the incident and jointly decide how to repair the harm that has occurred. In a further preventative step, they also consider what should be done differently in the future if a conflict arises again ([Costello et al., 2009](#); [Gregory et al., 2015](#)). Finally, there is a general consensus that RP should be implemented as a whole-school approach, integrating proactive and reactive practices ([Kane et al., 2009](#); [Morrison & Vaandering, 2012](#)), whereby educators provide both high control and high support to students. The overarching philosophy is characterised by doing things *with* people, rather than *to* them or *for* them, and therefore particular care should be taken to avoid embarrassment ([Wachtel, 2013](#)).

The appeal of RP lies in its potential to offer an effective alternative to behaviour management systems or exclusionary disciplinary systems, such as zero tolerance policy ([Morrison & Vaandering, 2012](#)). Despite documentation that RP is used globally, for example in Australia, Brazil, Canada, England, Hong Kong, New Zealand, Scotland, South Africa and the United States ([Anfara et al., 2013](#); [Grossi & dos Santos, 2012](#); [Wong, Cheng, Ngan & Ma, 2011](#)), the quality of research evaluating its efficacy in reducing conflicts in schools appears rather low ([Acosta, Chinman, Ebener, Phillips, Xenakis & Malone, 2016](#); [Green et al., 2019](#)). Correlational studies and descriptive reports suggest a positive influence of RP regarding student attendance ([González, 2012](#); [McMorris, Beckman, Shea, Baumgartner & Eggert, 2013](#)), students suspensions ([Jain, Bassey, Brown & Kalra, 2014](#); [Lewis, 2009](#); [McMorris et al., 2013](#); [Sumner, Silverman & Frampton, 2010](#)), student discipline ([González, 2012](#); [Lewis, 2009](#)), relationships in schools ([Jain et al., 2014](#); [Lewis, 2009](#); [McCluskey et al., 2008](#)), and school climate ([Jain et al., 2014](#); [Mirsky, 2007](#)). However, the methodological quality of studies varies widely and many studies are based only on qualitative surveys with a single measurement point, or lack comparison groups or randomised allocation (e.g. [McCluskey et al., 2008](#)). Furthermore, descriptions of the methodological procedures for the implementation of RP or data collection can be vague ([González, 2012](#); [Mirsky, 2007](#); [Sumner et al., 2010](#)). These methodological concerns raise questions about the validity and generalisability of the findings.

Mapping the effectiveness of RP in schools is thus crucial before implementing RP more widely. To our knowledge, no studies have systematically examined the extent of quantitative research on this topic. Previous reviews examined the effectiveness of RP in schools, but tended to focus on specific countries ([Fronius, Darling-Hammond, Persson, Guckenberg, Hurley & Petrosino, 2019](#)) or specific outcomes (e.g. disparities in school discipline in [Kline, 2016](#)), or primarily included qualitative

**Table 2**

Search Strings for all databases in the systematic review.

|   |  |
|---|--|
| 1 | "restorative approach*" OR "restorative justice*" OR "restorative practice*" |
| 2 | "school*" OR "teach*" OR "student*" OR "education*" OR "classroom*"          |
| 3 | #1 AND #2  |

findings ([Anfara et al., 2013](#)). Considering the identified gaps in current knowledge, the research question for this systematic review was: What is the effectiveness of RP in school environments in reducing conflicts? Therefore, we are primarily interested in any assessment of student-related conflicts. Furthermore, we are interested in measures relating to interpersonal relationships at school, school climate, and personal factors as secondary outcomes.

## 2. Method

The objective of this review was to examine potential evidence for the effectiveness of RP in reducing conflicts in school environments. In conducting and reporting this systematic review we adhere to the PRISMA guidelines ([Moher, Liberati, Tetzlaff, Altman & PRISMA Group, 2019](#)).

### 2.1. Search strategy

A systematic search of the literature was conducted in December 2019 to identify all studies reporting the effectiveness based on quantitative research of RP in school environments in reducing conflicts. We systematically searched abstracts in the following electronic databases: PsycINFO, PsycARTICLES, ASSIA, Social Science Citation Index, and ERIC, using the search terms "restorative approach\*", "restorative justice\*", "restorative practice\*", "school\*", "teach\*", "student\*", "education\*" and "classroom\*" which were combined using Boolean operators and truncation marks ([Table 2](#)). Using the ancestry method, reference lists and citation records of all included studies were screened to identify additional papers that may not have fulfilled the search term criteria.

### 2.2. Study selection

The review eligibility criteria were as follows: studies had to measure the outcome of RP in school settings, whereby (1) participants could be students or teachers from primary, middle or secondary schools, (2) study designs delivered quantitative outcome data, and (3) the study

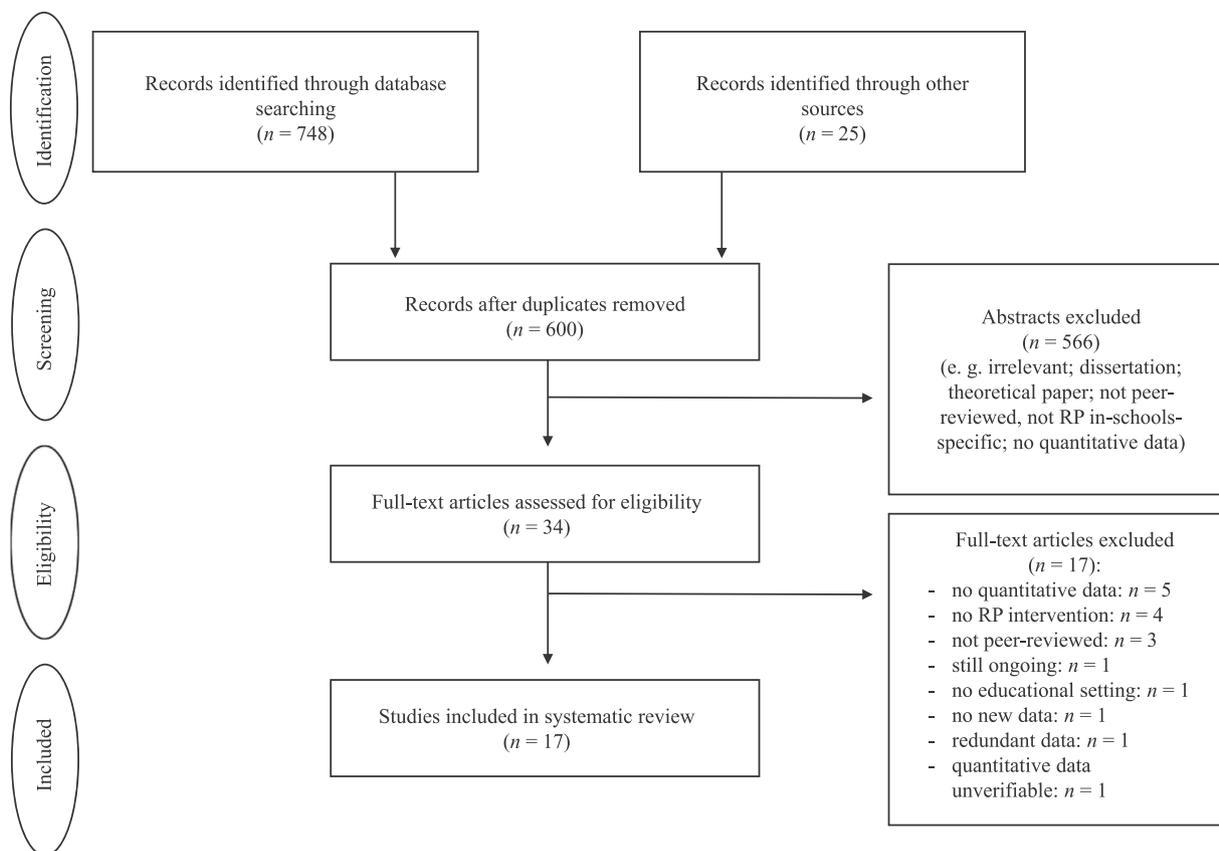


Fig. 1. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) diagram illustrating study selection.

had undergone peer-review. Therefore, studies from randomised controlled trials (RCT) down to case studies were included. No limits or restrictions were applied regarding publication date or language. Exclusion criteria concerned (1) studies relating to university teaching or other educational settings not covered in the inclusion criteria, as well as studies relating to RP in a judicial context, (2) publication type (theses and dissertations), and (3) non-intervention studies (e.g. commentary, reflections, narrative review, selective review, theoretical paper). Fig. 1 presents a flow diagram of the study selection process.

Study selection was performed by the first author, with twenty per cent of records randomly selected and screened by another reviewer. After reading abstracts, reviewer agreement was 99.2% with a kappa coefficient of 0.99 for studies to proceed to full-text evaluation with all eligibility criteria taking into account. The overall inclusion rate for the screening of titles and abstracts was 5.7%. Full-text review of all articles which were assessed for eligibility ( $n = 34$ ) was performed independently by the two reviewers, which resulted in a kappa coefficient of 0.94 and an agreement of 97.1% for inclusion in the review. Again, all eligibility criteria were taken into account. Disagreements were resolved through consultation with a third reviewer.

### 2.3. Working definitions

For this systematic review, we consider RP that include a way of working with students which focuses on reducing conflicts in interpersonal interaction (Hendry, 2009). We define a conflict as an interpersonal disagreement between people based on their ideas, needs or beliefs (Hendry, 2009). In this regard, RP comprises the use of both informal and formal methods. Some methods proactively build relationships and community to prevent conflict and wrongdoing, and some methods are used reactively when harm has already occurred (Wachtel, 2013).

### 2.4. Outcomes

Primary outcomes consider assessments of student-related conflicts in school settings. This may include direct measures of conflict, such as bullying and aggression, as well as the consequences of conflict situations at school, such as expulsion or suspension rates.

Secondary outcomes of interest include measures relating to interpersonal relationships at school, school climate, and personal factors. We are thereby particularly interested in examining the breadth of outcomes that have been assessed in studies implementing RP.

### 2.5. Data extraction

The following data were extracted for included studies: author and year, journal name, title, country, study population, study design, implementation of RP, outcomes, study aims and findings.

### 2.6. Collating, summarising and reporting the results

First, we summarised the nature, extent and characteristics of selected studies. Second, we synthesised individual study findings guided by our primary and secondary outcomes. A standardised quality assessment tool, the Effective Public Health Practice Project Quality Assessment Tool (EPHPP Tool) for Quantitative Studies (National Collaborating Centre for Methods & Tools, 2008), was used for critical appraisal of included studies and includes an assessment of the risk of bias within and across studies. We chose this tool because it enables the evaluation of different study designs using a single measure. Each criterion was rated as strong, moderate or weak, and then summed to obtain an overall score for each study.

### 3. Results

The search strategy identified 600 records after removal of duplicates, with 34 studies proceeding to full text review. Main reasons for excluding studies were: not referring to RP in the school setting, not reporting quantitative data, or not having undergone peer review. Seventeen studies met all review eligibility data. The study selection process is illustrated in Fig. 1. Table 3 provides summary information of each of the included studies.

#### 3.1. Publication characteristics

The included studies were published between 2002 and 2019, with almost half (47.1%,  $n = 8$ ) of the included studies published since 2018. Studies were predominantly conducted in the USA ( $n = 13$ ), with other studies conducted in the UK ( $n = 2$ ), Australia and Hong Kong (both  $n = 1$ ).

The studies represented a wide range of disciplines, with studies being published in journals reporting on topics of children, youth and adolescents, education, educational administration, psychology, medicine, criminology, violence, restorative justice, and conflict resolution. Three studies were not published in journals but equally underwent peer review. The studies of McCold (2008) and Gregory & Clawson (2016) were published in books in the field of sociology, criminology and school discipline, whereas the report by Augustine et al. (2018) was published by the non-profit, nonpartisan RAND Corporation.

#### 3.2. Study design

Most of the included studies were correlative studies in a cross-sectional design ( $n = 6$ ). Rainbolt, Fowler and Mansfield (2019), and Brown (2017) chose a one-group post-test-only design, while Stinchcomb, Bazemore and Riestenberg (2006), Mansfield, Fowler and Rainbolt (2018), and Morrison (2002) chose a one-group follow-up design. Norris (2019) and Wong et al. (2011) conducted a quasi-experimental pre-post design and compared three groups with different treatments. Norris (2019) examined the impact of a traditional RP whole-school model, a reactive-only model and a preventative-only whole-school model on student outcomes, whereas Wong et al. (2011) screened schools for indicators of implementation of a RP whole-school approach after intervention and scheduled the schools into three different degrees of RP implementation: a traditional RP whole-school approach, a partial RP whole-school approach and a control group. Hashim, Strunk and Dhaliwal (2018) conducted an interrupted time series analysis to examine the effect of school discipline reforms, including RP. Acosta, Chinman, Ebener, Malone, Phillips, and Wilks (2019), Augustine and colleagues (2018) as well as Bonell and colleagues (2018) conducted cluster RCTs.

Most studies did not provide information on the criteria used to select the participating schools and did not employ random sampling. Three of the included studies (17.6%) explicitly reported randomised group allocation. The statistical evaluation of the outcomes also differed greatly from one another, and at a minimum provided only reporting of frequencies without statistical analysis.

#### 3.3. Implementation of RP

The degree of detail in the description of the training varied widely. Furthermore, the time between implementation and post-intervention outcome assessment varied as well from six days to three years. Three studies assessed outcomes three years post implementation of RP (Bonell et al., 2018; Rainbolt et al., 2019; Stinchcomb et al., 2006); however, most studies ( $n = 9$ ) had a duration of between one and two years (Acosta et al., 2019; Augustine et al., 2018; Gregory & Clawson, 2016; Gregory et al., 2015; Hashim et al., 2018; Mansfield et al., 2018; Morrison, 2002; Norris, 2019; Wong et al., 2011). For another

five studies, no information was given about the time between implementation and data/outcome assessment.

Regarding the description of the elements of RP that were implemented, three different types of implementation could be identified. Four studies provided a description of their implementation plan in line with the guidelines from the International Institute of Restorative Practices (IIRP). The IIRP, a private company headquartered in Pennsylvania, developed the SaferSanerSchools™ Whole-School Change program, which is a two-year implementation programme intended to engage all school staff in RP (International Institute for Restorative Practices, 2011). Most authors described the training in RP in detail ( $n = 10$ ), some studies stated that different RP were implemented in schools without specifying this more precisely ( $n = 3$ ). Some described the elements of RP that have been implemented ( $n = 12$ ), such as affective statements, restorative questions, circles, conferences etc. However, in multiple studies it remained unclear to what extent specific methods of RP had been implemented or if only the philosophy of RP had been established. In two studies, elements of RP were part of a training programme comprising elements from other interventions: the Learning Together intervention and the Responsible Citizenship Program (Bonell et al., 2018; Morrison, 2002).

#### 3.4. Study samples

Participant demographics also varied between studies. The majority ( $n = 12$ ) of the included studies focused on students, with two studies focusing on teachers (Brown, 2017; Rainbolt et al., 2019), or including both students and teachers (Augustine et al., 2018; Gregory & Clawson, 2016; Gregory et al., 2015). The same sub-sample of 29 teachers was included in two different studies (Gregory & Clawson, 2016; Gregory et al., 2015); however, the respective data analyses related to different research questions. Meanwhile, three studies included data from different school years in the same school district (Anyon et al., 2014; 2016; Gregory, Huang, Anyon, Greer & Downing, 2018) and two studies included data from the same school in different years (Mansfield et al., 2018; Rainbolt et al., 2019).

School grades under investigation included all levels eligible for inclusion in this review: elementary school, secondary school, high school and grade K-12. Some studies focused on individual grades (Bonell et al., 2018), whereas other studies included any grade in their analysis (Gregory et al., 2018). One study referred to a special school focused on troubled and at-risk youth (McCold, 2008).

#### 3.5. Primary outcomes

We identified two types of primary outcomes assessing student-related conflicts: outcomes on the individual student level, such as behavioural referrals and suspension rates, and on the interpersonal level, such as interpersonal aggression and bullying.

##### 3.5.1. Individual student outcomes: behaviour records/referrals, suspensions and expulsions

Quantitative data regarding measures of conflicts at school and their consequences generally concerned behaviour records/referrals, suspensions and expulsions. Of the nine studies providing data concerning our primary outcome, five adopted a correlational design, two used pre-post designs, one conducted an interrupted time series analysis, and only a single RCT was conducted.

Starting with the more robust study design, the cluster RCT by Augustine and colleagues (2018) found larger overall decreases in the number of suspension days and suspension rates for elementary students in the intervention group than in the control group, but did not report any change in outcomes concerning the rate of arrests.

Using time series analysis, Hashim and colleagues (2018) examined the impact of school discipline policy reforms over a time span about 10 years. They found a continual decrease in suspensions with a link to

**Table 3**  
Study characteristics of the studies included in this systematic review ( $n = 17$ ).

| First author, year  | Study Design                                | Study Population   | Intervention  | Primary outcomes  | Secondary outcomes   |
|---------------------|---|--|---|---|--|
| Acosta et al., 2019 | Cluster randomised controlled trial         | $N = 2771$ middle school students (EG: $n = 977$ ; CG: $n = 1794$ )<br>$n = 13$ schools<br>Age: 11–12 years<br>Sex: EG/CG: 48%/ 50% female<br>Grade: 6–7<br>Country: USA               | EG: IIRP's SaferSanerSchools™ Whole-School Change program: training of all school staff in all elements of RP<br>CG: no intervention, standard practice<br>Duration: 2 years  | <b>Bullying victimisation:</b><br>- No between-group differences for either physical bullying ( $OR = 1.18$ ), emotional bullying ( $OR = 1.06$ ), or cyberbullying ( $OR = 0.89$ )<br>- Students' self-reported experience with RP was associated with reduced physical ( $OR = 0.74, p < .05$ ) and cyberbullying ( $OR = 0.69, p < .05$ ), but not emotional bullying ( $OR = 0.82, ns$ )  | <b>School climate, school connectedness, peer attachment &amp; social skills:</b><br>- No improvements for students in intervention schools on either outcome ( $p$ 's $> 0.14$ , median effect size estimate: $d = 0.57$ )<br>- Students' self-reported experience with RP was associated with improved school climate and connectedness, peer attachment, and social skills (median semi-partial standardised regression coefficient $\beta = 0.42, p < .05$ ) |
| Anyon et al., 2014  | Correlational study, cross-sectional design | $N = 87,997$ students<br>$n = 183$ schools<br>Age: not reported<br>Sex: 51% female<br>Grade: K-12<br>Country: USA  | 2 types of voluntary staff training in 'restorative interventions': 4 h introductory training in prevention (e.g. community-building circles) and/or 2 days training in the philosophy, origin, effectiveness, proactive and reactive methods of RP | <b>Suspension:</b><br>- Reduced odds of out-of-school suspension for students who participated in a restorative intervention (circles, mediations, or conferences) to resolving their discipline problems ( $OR = 0.73, p < .01$ ) or an in-school suspension ( $OR = 0.37, p < .001$ ), when controlled for severity of referral reason and demographic covariates   | <b>Expulsion &amp; law enforcement:</b><br>- Alternative approaches to resolving discipline problems did not protect students from referral to law enforcement (behaviour contracts: $OR = 2.3, p < .001$ , in-school suspension: $OR = 0.85, p > .05$ , RP: $OR = 1.08, p > .05$ ) or from expulsion (behaviour contracts: $OR = 1.56$ , in-school suspension: $OR = 0.79$ , RP: $OR = 1.77, p$ 's $> 0.05$ )<br>None   |
| Anyon et al., 2016  | Correlational study                         | $N = 9921$ students, who were issued at least one Office Disciplinary Referral (ODR)<br>$n = 180$ schools<br>Age: not reported<br>Sex: $OR = 1.04$ male<br>Grade: K-12<br>Country: USA | 2 types of voluntary staff training in 'restorative interventions': 4 h introductory training in prevention (e.g. community-building circles) and/or 2 days training in the philosophy, origin, effectiveness, proactive and reactive methods of RP | <b>Suspension &amp; office discipline referrals (ODR):</b><br>- Reduced odds of suspension and ODR in the second semester for students who received RP (circles, mediations, or conferences) as consequences for referrals in the first semester (suspension: $OR = 0.07, p < .001$ , ODR: $OR = 0.21, p < .001$ ), when controlled for reasons for and frequency of ODR and a range of school and student characteristics<br>- Lower odds of receiving a second-semester ODR in schools for RP participants that had higher schoolwide RP rates ( $OR = 0.86, p < .001$ )<br>- Negative correlation between the school-level RP rate and receiving suspensions for all students ( $OR = 0.13, p < .01$ ) | None   |

(continued on next page)

Table 3 (continued)

| First author, year     | Study Design                        | Study Population  | Intervention  | Primary outcomes  | Secondary outcomes  |
|------------------------|-------------------------------------|---|---|---|---|
| Augustine et al., 2018 | Cluster randomised controlled trial | <p><i>N</i> for students not reported<br/> <i>n</i> = 460 teachers<br/> <i>n</i> = 650 staff members (group membership of teachers and staff members not reported)<br/> <i>n</i> = 44 schools (EG &amp; CG; <i>n</i> = 22)<br/>           Age: not reported<br/>           Sex: not reported<br/>           Grade: K-12<br/>           Country: USA</p> | <p>EG: IIRP's SaferSanerSchools™ Whole-School Change program, two-years implementation programme, including at least two and voluntarily four days of professional development relating to RP for all school staff; coaching, restorative leadership teams and learning groups<br/>           CG: business as usual</p> | <p><i>Suspension rates:</i><br/>           - Between-group comparison: 16% reduction in days lost to suspensions, (<math>p &lt; .05</math>, impact estimate: <math>-.06</math>). Within group reductions over 2 years: EG = 36%, CG = 18%<br/>           - Decreased suspension rates only for elementary students in EG<br/>           - No differences in suspension rates for middle school students</p> | <p><i>Student-teacher relationships:</i><br/>           - 63% of teachers reported moderate to great improvements<br/> <i>Students' perceptions on classroom climate:</i><br/>           - No between-group difference, but lower classroom climate ratings by students in EG<br/> <i>Academic outcomes:</i><br/>           - Increased PSAT scores for 10th grade students (<math>p &lt; .01</math>, impact estimate: 0.118).<br/>           - Reduced elementary and middle school math performance (<math>p &lt; .05</math>, impact estimate: 0.068), and reduced academic performance of Black students in elementary and middle schools (<math>p &lt; .01</math>, impact estimate: <math>-.15</math>) in EG<br/> <i>Teachers' perceptions of school climate &amp; working conditions due to the implementation of RP:</i><br/>           - RP was associated with increases in overall teaching and learning conditions (<math>p &lt; .05</math>, impact estimate: 0.331), teacher leadership (<math>p &lt; .05</math>, impact estimate: 0.331), school leadership (<math>p &lt; .01</math>, impact estimate: 0.292), conduct management (<math>p &lt; .05</math>, impact estimate: 0.308) and work safety (<math>p &lt; .05</math>, impact estimate: 0.144)<br/> <i>Absenteeism:</i><br/>           - No reduction in absenteeism for EG in second year of the intervention<br/>           - Between-group differences in attendance rates for elementary students (<math>p &lt; .05</math>, impact estimate: 0.082) and students in the individualised education programme (<math>p &lt; .01</math>, impact estimate: 0.149) in favour of the EG, but not in other K-12 groups<br/> <i>Mobility:</i><br/>           - No link between RP implementation and student mobility (changing schools)<br/> <i>Arrests:</i><br/>           - No difference between the groups</p> |

(continued on next page)

Table 3 (continued)

| First author, year       | Study Design                                | Study Population  | Intervention  | Primary outcomes  | Secondary outcomes  |
|--------------------------|---|---|---|---|---|
| Bonell et al., 2018      | Cluster randomised controlled trial         | <i>N</i> = 6667 students (EG: <i>n</i> = 3320, CG: <i>n</i> = 3347)<br><i>n</i> = 40 schools (EG & CG: <i>n</i> = 20)<br>Age: <i>M</i> = 12 ( <i>SD</i> = 0.4) years<br>Sex: EG/CG: 55.2%/ 50.2% female<br>Grade: 7<br>Country: England | EG: 'Learning Together' intervention: all staff were trained in RP, in-depth training for selected staff, manual-guided action group meetings between students and teachers twice per term, 5–10 h per year of lessons on social and emotional skills for students in years 8–10, implementation of proactive and reactive RP<br>CG: standard practice<br>Duration: 3 years                                     | <i>Bullying victimisation:</i><br>- Between group-difference: higher bullying score at 36 months in the CG (adjusted difference: -.03, <i>p</i> = .04, adjusted effect size: -.08)<br><i>Perpetration of aggression:</i><br>- No between-group difference at 36 months (adjusted difference: -.13, <i>p</i> = .42, adjusted effect size: -.03)  | <i>Quality of life &amp; psychological wellbeing:</i><br>- Between-group difference at 36 months in favour of EG (quality of life: adjusted difference: 1.44, <i>p</i> < .001, adjusted effect size: 0.14; total wellbeing: adjusted difference: 0.33, <i>p</i> < .05, adjusted effect size: 0.07)<br><i>Emotional problems, conduct problems, hyperactivity &amp; peer problems:</i><br>- No between-group difference at 36 months, but tendencies in favour of EG<br><i>Consumer &amp; risk behaviour:</i><br>- Lower odds in EG of having ever smoked regularly (adjusted risk difference: -.03, <i>p</i> < .001), having ever drunk alcohol (adjusted risk difference: -.03, <i>p</i> < .01), and lower odds of having ever been offered or tried illicit drugs (adjusted OR: 0.51, <i>p</i> < .01) or having ever been in contact with police in the past 12 months (adjusted risk difference: -.02, <i>p</i> < .05)<br><i>Listening to each other:</i><br>- Strong agreement about a positive listening culture in school (directional lean <sup>b</sup> : <i>p</i> < .001)<br><i>Students' voice:</i><br>- Strong agreement about a positive perceived voice and empowerment of students (directional lean: <i>p</i> < .001)<br><i>Quality of teacher-student-relationship:</i><br>- No link between teacher-reported RP implementation and teacher-student relationships ( $\beta$ = -.05, <i>ns</i> )<br>- Link between higher student-reported RP and greater teacher respect ( <i>r</i> = 0.58, $\beta$ = 0.12, <i>p</i> < .01)<br>- Link between teacher-reported cooperation of students and student-reported respect of teachers ( $\beta$ = 0.20, <i>p</i> < .01)<br>- No link between racial group membership and teacher respect |
| Brown, 2017 <sup>a</sup> | One-group post-test-only design             | <i>N</i> = 60 staff members<br><i>n</i> = 2 middle schools<br>Age: not reported<br>Sex: not reported<br>Grades: 6–8<br>Country: USA   | Participation of students in community building circles during advisory periods   | None  |   |
| Gregory et al., 2015     | Correlational study, cross-sectional design | <i>N</i> = 412 students<br><i>n</i> = 29 teachers<br><i>n</i> = 2 high schools<br><br>Age: not reported<br>Sex: 47% female students, almost three-quarter female teachers<br>Country: USA   | IIRP's SaferSanerSchools™<br>Whole-School Change program, two day workshops with teachers, administrators, and staff, covering the following elements of RP: affective statements, fair process, management of shame, proactive circles, restorative questions and reactive circle. IIRP trainers provided two full days of consultation in each school including observations, modelling and targeted planning | <i>School discipline records:</i><br>- No link between teacher-reported RP implementation and discipline referrals<br>- Student-reported RP implementation predicted misconduct/defiance referrals issued to Latino and African American students ( $\beta$ = -.44, <i>p</i> < .05), but not for referrals issued to Asian and White students ( $\beta$ = -.34, <i>p</i> = .10)<br>- Difference in referrals between Asian/White and Latino/African American students, which was smaller for student-reported high implementing RP teachers (low RP implementing teachers: <i>t</i> (15) = 3.21, <i>p</i> = .006, <i>r</i> = 0.99, high RP implementing teachers: <i>t</i> (12) = 2.69, <i>p</i> < .02, <i>r</i> = 0.613) |   |

(continued on next page)

Table 3 (continued)

| First author, year      | Study Design                                | Study Population  | Intervention  | Primary outcomes  | Secondary outcomes |
|-------------------------|---|---|---|---|--------------------|
| Gregory & Clawson, 2016 | Correlational study                         | <i>N</i> = 412 students<br><i>n</i> = 29 teachers<br><i>n</i> = 2 high schools<br><br>Age: not reported<br>Sex: 47% female students, almost three-quarter female teachers<br>Country: USA | IIRP's SaferSanerSchools™<br>Whole-School Change program, two day workshops with teachers, administrators, and staff, covering the following elements of RP: affective statements, fair process, management of shame, proactive circles, restorative questions and reactive circle. IIRP trainers provided two full days of consultation in each school including observations, modelling and targeted planning | <i>Discipline records:</i><br>- Link between RP and fewer misconduct/defiance referrals (decrease of 21% after second year during RP implementation)<br>- Overall reductions regarding discipline records, but racial/ethnic and gender disparities persisted (e.g. 11% of White males vs. 43% of African American males vs. 32% of Latino males received one or more misconduct/defiance referral)<br>- Link between teachers who were student-reported frequently using affective statements and fewer misconduct/defiance office disciplinary referrals (ODRs) to African American and Latino males ( $\beta = -.57, p < .05$ ) and females ( $\beta = -.60, p < .04$ )<br>- No link between any RP element and ODRs issued to White students  | None               |
| Gregory et al., 2018    | Correlational study, cross-sectional-design | <i>N</i> = 9039 discipline-referred students<br><i>n</i> = 193 schools<br>Age: not reported<br>Sex: 31.8% female<br>Grade: K-12<br>Country: USA   | Two types of training on a voluntary basis: a half-day session on preventive RP like peace circles and affective statements and a 2-day session on the use of reactive RP (conferences and mediations) to address discipline incidents  | <i>Out-of-school suspension (OSS):</i><br>- Link between disciplinary consequences (RP, in-school suspension (ISS), and behavioural contracts) and OSS ( $p$ 's < 0.001)<br>- Almost 50% increase in the odds of an OSS when receiving at least one behavioural contract ( $OR = 1.96, p < .001, d = 0.40$ )<br>- 85% decrease in the odds of an OSS when receiving at least one ISS ( $OR = 0.15, p < .001, d = 1.05$ ), while controlling for all other variables<br>- 69% decrease in the odds of an OSS when receiving one or more restorative intervention, while controlling for all other variables ( $OR = 0.31, p < .001, d = 0.65$ )<br>- Students in schools that responded to a greater percentage of infractions with RP had a lower likelihood of suspension ( $OR = 0.98, p < .001$ )<br>- The benefits of RP were relatively similar across racial groups | None               |
| Hashim et al., 2018     | Interrupted time series analysis            | <i>N</i> = administrative data of 1.44 million students<br><i>n</i> = 785 schools<br>Age: not reported<br>Sex: not reported<br>Grade: K-12<br>Country: USA                                | School discipline reform: positive behaviour intervention support (since year 2006–2007), suspension bans (since year 2011–2012), training in RP and implementation of RP (since year 2014–2015)  | <i>Suspension rates:</i><br>- Decrease of suspension rate: starting at a rate of 7.1% in 2004–2005 and then dropping at the rate of 0.60 percentage points per year, link between steep decreases with the implementation of suspension bans (2012–2013: $R^2 = 0.20, -.016, p < .001$ ) and RP (2014–2015: $R^2 = 0.20, -.010, p < .001$ )<br>- Link between steeper decreases of frequently disciplined subgroups and RP (e.g. male: $R^2 = 0.20, -.008, p < .001$ , special educational needs (SEN) students: $R^2 = 0.20, -.008, p < .001$ ), although suspension gaps between Black and non-Black students, and between SEN and non-SEN students persisted   | None               |

(continued on next page)

Table 3 (continued)

| First author, year                  | Study Design  | Study Population  | Intervention   | Primary outcomes  | Secondary outcomes  |
|-------------------------------------|---|---|--|---|---|
| Mansfield et al., 2018 <sup>a</sup> | One-group pre-post design, descriptive study                                  | N = about 1400 students<br>n = 1 high school<br>Age: not reported<br>Sex: not reported<br>Grades: 9–12<br>Country: USA  | Implementation of written restorative questions when serving in-school detention (2008), restorative conferences (since 2010–2011), training in RP for interested teachers (since 2013), classroom implementation of RP (since 2013), implementation of 'behavioural Interventions and Supports®' (since 2010) | <i>Suspensions:</i><br>- Decrease in frequency of suspensions (2010–2011: 19% received at least 1 day of in-school suspension (ISS) and 12% at least 1 day of out-of-school suspension (OSS); 2014–2015: 7% at least one ISS and 7% at least one OSS)<br>- Decrease in OSS for White (from 23% to 12%) and Black students (from 7% to 4%) from 2010–2011 to 2014–2015<br>- Decrease in suspensions of students with disabilities, while still at least 10% higher suspension rates compared to total students<br>- Decrease in suspension rates for boys and girls, with boys being suspended more often than girls (2014–2015: 9% boys, 4% girls)<br>- Decrease of 66% and almost 50% in number of students who received ISS and OSS more than once in a given school year from 2010–2011 to 2014–2015 | None  |
| McCold, 2008                        | Cross-sectional study   | N = 1636 discharged students from CSF Buxmont schools<br>n = 8 school programmes<br>Age: 11–19 years, average: 15.5 years<br>Sex: 30% female<br>Grade: not reported<br>Country: USA | CSF Buxmont schools are non-secure alternative schools and community treatment settings for adjudicated delinquent and at-risk youth, which utilise RP, especially restorative conferences   | None  | <i>Post-release recidivism rates (juvenile &amp; adult court records):</i><br>- Link between recidivism rates and length of programme participation ( $r = -.11, p < .001$ ), this relationship held after controlling for some risk factors such as age, gender, race, and prior offending; however, the statistical models indicate that there is a possibility for another causally unknown prior factor accounting for the relationship between recidivism and the length of participation<br><i>Shame management:</i><br>- Small increase in students' reported use of adaptive shame management skills (from 83% to 87%)<br>- Decrease in the use of maladaptive shame management skills (decrease in feelings of rejection by others and displacement of wrongdoing onto others from 33% to 20%)<br>- Decrease in student-reported shame displacement strategies (from 27% to 13%)<br><i>Feelings of safety within the school community:</i><br>- Increase of students' feelings of safety within the school (from 2.9 to 3.8 on a four-point scale) |
| Morrison, 2002 <sup>a</sup>         | One-group pre-post-design obscure, whether within- or between-subjects-design | N = 30 students<br>n = one primary school<br>Age: 10–11 years<br>Sex: not reported<br>Grade: 5<br>Country: Australia  | 'Responsible Citizenship Program', which is grounded in principles of restorative justice (community building and conflict resolution)<br>Duration: students participated for 5 weeks (10 h in total)  | None  |   |

(continued on next page)

Table 3 (continued)

| First author, year                 | Study Design  | Study Population   | Intervention  | Primary outcomes | Secondary outcomes   |
|------------------------------------|---|--|---|------------------|--|
| Norris, 2019                       | Quasi-experimental study, School 1: pre-post-design School 2 & 3: linked cross-sectional design with 2 × 2 between subject analysis | n = 3 schools<br>School 1:<br>n = 19 'referred cases'<br>Age: not reported<br>Sex: 21% female<br>School 2 & 3:<br>Grade: not reported<br>School 2:<br>n = 181 students<br>Age: not reported<br>Sex: not reported<br>Grade: 7, 9 & 11<br>School 3:<br>n = 276 students<br>Age: not reported<br>Sex: not reported<br>Grade: 7, 9 & 11<br>Country: UK | School 1: 'reactive-only model of RP' consisting of initial meetings, restorative conferences, and follow-up meetings facilitated by a RA Officer<br>School 2: 'traditional whole-school approach', proactive and reactive practices<br>In school 1 and 2, the local Youth Offending Teams RP officers were based in the schools for two full days a week, both schools trained all staff for one day<br>School 3: 'proactive-only whole-school approach', training in proactive practices for all staff members, training with further instructions on a voluntarily basis | None             | <i>Happiness &amp; school engagement:</i><br>- School 1: No difference in mean happiness scores ( $t = 1.008, p = .327, r = 0.163$ ) and in school engagement scores ( $t = -.494, p = .627, r = -.036$ ) before and after participation in RP<br>- Between-group difference between School 2 and School 3 in happiness and school engagement:<br>- Happiness scores for School 2 were higher than that of School 3 at T1; at T2 School 3 improved their scores, whereas there was a decline in School 2 ( $F(2,00) = 14.72, p = .001, NE$ )<br>- Small difference between school engagement scores at T1 between School 2 and School 3; at T2 the scores of School 2 were decreased and School 3's scores were increased ( $F(2,00) = 5.82, p = .02, NE$ )<br><i>RP implementation:</i><br>- Around 69% of teachers used RP "often" or "almost daily", nearly 31% "seldom" or "never"<br>- 56% observed others using RP "often" or "almost daily", 44% of respondents indicated seeing RP used "seldom"<br>- 25% believed that "most" students participate in elements of RP, whereas 75% indicated only "some" students were participating in RP so far<br>- 24 of 33 respondents selected "Affective Statements" as being most effective, according to the 11 elements of the SaferSanerSchools™ model<br>- Participants rated that RP "often" or "seldom" curb student misbehaviour<br><i>Relationships</i><br>- Participants rated student-teacher relationships as being respectful, moreover they chose student-teacher-relationships as being the most respectful compared to staff- and student-student-relationships |
| Rainbolt et al., 2019 <sup>a</sup> | One-group post-test-only design   | N = 43 teachers<br>n = 1 high school<br>Age: not reported<br>Sex: not reported<br>Grades: 9–12<br>Country: USA   | Implementation of written restorative questions when serving in-school detention (since 2008), restorative conferences (since 2010–2011), training in RP for interested teachers (since 2013), classroom implementation of RP (since 2013), implementation of 'behavioural Interventions and Supports©' (since 2010)  | None             |  |

(continued on next page)

Table 3 (continued)

| First author, year                   | Study Design  | Study Population  | Intervention   | Primary outcomes   | Secondary outcomes  |
|--------------------------------------|---|---|--|--|---|
| Stinchcomb et al., 2006 <sup>a</sup> | One-group pre-post-design, observational study                                    | <i>N</i> for students not reported<br><i>n</i> = 3 schools (2 high schools, 1 elementary school)<br>Age: not reported<br>Sex: not reported<br>Grade: 7 & 8 from one school and 2 whole elementary schools<br>Country: USA   | Basic training on RP for all school staff and positive behaviour support practices. Opportunity for an intensive training on RP (mostly on circles). RP planners are located in each of the schools, conducting restorative and community-building circles and they promote 'Make the Peace', a state-wide campaign to promote alternatives to violence<br>Duration: about 3 years | <i>Out-of-school suspensions (OSS)</i> :<br>- Decrease in the frequency of OSS in every school (1998–1999: 30, 27 and 110 OSS compared to 2000–2001: 11, 4 and 55 OSS)<br><i>In-school suspensions (ISS)</i> :<br>- Decrease of ISS in one elementary school (1999–2000: 126 ISS and 2000–2001: 42 ISS)<br>- Increase of ISS in another elementary school (1998–1999: 8 ISS and 2000–2001: 36 ISS; simultaneously there was a policy change designed to keep students in school rather than giving students a free day); in another high school ISS was not an option<br><i>Behaviour referrals</i> :<br>- Behaviour referrals were only recorded in the two primary schools<br>- Decrease in one school (1998–1999: 1143 and 2000–2001: 407)<br>- Increase in the other school (1998–1999: 361 and 2000–2001: 459)<br><i>Bullying</i> :<br>- Decrease in overall bullying in the full implementing EG ( $t = 3.41, p < .001, ES = 0.18$ ) and partial implementing EG ( $t = 2.40, p < .05, ES = 0.10$ )<br>- Increase in overall bullying in the CG ( $t = -.25, p < .01, ES = 0.19$ )<br>- Decrease in physical ( $t = 3.05, p < .01, ES = 0.16$ ) and exclusion bullying ( $t = 4.48, p < .001, ES = 0.24$ ) in the full implementing EG<br>- Decrease in exclusion bullying ( $t = 5.33, p < .001, ES = 0.22$ ) in the partial implementing EG<br>- Increase in physical ( $t = -2.26, p < .05, ES = 0.16$ ), verbal ( $t = -3.20, p < .05, ES = 0.22$ ) and extortion bullying ( $t = -2.55, p < .05, ES = 0.18$ ) in the CG | <i>Attendance</i> :<br>- Improved average daily attendance in one elementary school from 85% during 1997–1998 to 95.5% in 2000–2001, as noted by administrators<br><i>Expulsions</i> :<br>- Only one expulsion in all schools in the first year of implementation   |
| Wong et al., 2011                    | Quasi-experimental study, 2-year longitudinal with pre-post within-subject design | <i>N</i> = 1480 students<br><i>n</i> = 4 secondary schools (EG: 'full implementation': <i>n</i> = 1 school, 'partial implementation': <i>n</i> = 2 schools, CG: <i>n</i> = 1 school)<br>Age: 85% between 12 and 14 years<br>Sex: about 50% female<br>Grade: 7–9<br>Country: Hong Kong | Two types of EG: full RP implementing schools and partial implementing schools; training in a RP whole-school programme, training in school harmony programmes<br>CG: no training and implementation of RP<br>Duration: full implementation of RP programme lasted for 15 months   | <i>Bullying</i> :<br>- Decrease in overall bullying in the full implementing EG ( $t = 3.41, p < .001, ES = 0.18$ ) and partial implementing EG ( $t = 2.40, p < .05, ES = 0.10$ )<br>- Increase in overall bullying in the CG ( $t = -.25, p < .01, ES = 0.19$ )<br>- Decrease in physical ( $t = 3.05, p < .01, ES = 0.16$ ) and exclusion bullying ( $t = 4.48, p < .001, ES = 0.24$ ) in the full implementing EG<br>- Decrease in exclusion bullying ( $t = 5.33, p < .001, ES = 0.22$ ) in the partial implementing EG<br>- Increase in physical ( $t = -2.26, p < .05, ES = 0.16$ ), verbal ( $t = -3.20, p < .05, ES = 0.22$ ) and extortion bullying ( $t = -2.55, p < .05, ES = 0.18$ ) in the CG  | <i>Lack of empathy</i> :<br>- Decrease in the full implementing EG ( $t = 2.35, p < .05, ES = 0.12$ )<br><i>Hurting others</i> :<br>- Decrease in partial implementing EG ( $t = -5.04, p < .001, ES = 0.21$ )<br><i>Caring behaviour</i> :<br>- Increase in partial implementing EG ( $t = 4.78, p < .001, ES = 0.20$ )<br>- Decrease in CG ( $t = -3.15, p < .01, ES = 0.22$ )<br><i>Self-esteem</i> :<br>- Increase in the full implementing EG ( $t = 4.0, p < .001, ES = 0.21$ )<br><i>Level of school harmony</i> :<br>- Decreases in partial implementing EG ( $t = -2.00, p < .05, ES = 0.08$ ) and CG ( $t = -4.99, p < .001, ES = 0.34$ )<br><i>Sense of belonging</i> :<br>- Decreases in partial implementing EG ( $t = -4.81, p < .001, ES = 0.20$ ) and CG ( $t = -5.0, p < .001, ES = 0.34$ )<br><i>Positive perception toward teachers</i> :<br>- Decreases in partial implementing EG ( $t = -6.98, p < .001, ES = 0.29$ ) and CG ( $t = -4.23, p < .001, ES = 0.30$ ) |

EG: experimental group, CG: control group, NE: effect size cannot be calculated, ES = effect size,  $ZD/\sqrt{n}$ , in which *ZD* is the standardised score of the pairwise difference between pre- and post-test, ns: not significant.

<sup>a</sup> No effect sizes can be calculated for these studies.

<sup>b</sup> Survey responses were analysed using a Test of Symmetry and Disagreement.

new policies, like suspension bans or the implementation of RP. Moreover, they found an association between RP and steeper decreases of frequently disciplined subgroups, like male students or students with special educational needs.

Stinchcomb and colleagues (2006), and Mansfield and colleagues (2018) both used a one-group pre-post-test design and observed a decrease in suspensions following the implementation of RP, where by Stinchcomb and colleagues (2006) also focused on behavioural referrals and found mixed results.

Using a large sample of 87,997 students, Anyon and colleagues (2014) examined the association between specific risk and protective factors, including RP with the odds of out-of-school suspensions. Their multilevel logistic regression model showed that students had lower odds of out-of-school suspension if they participated in RP (circles, mediations or conferences) or in an in-school suspension to resolve their discipline problems, even when student covariates, such as severity of referral reason and demographics, were controlled for. In comparison, working with behaviour contracts did not reduce the odds of out-of-school suspension.

Using different data in the same school district, Anyon and colleagues (2016) also examined whether a first-semester participation in RP was associated with lower odds of receiving office discipline referrals and suspensions in the second semester. Whilst their findings confirmed these hypotheses, analysis of ethnicity covariates suggested that the suspension gap between Black and White students persisted through the use of RP.

Following this and within the same school district, Gregory and colleagues (2018) examined whether alternatives to suspensions, such as (whole-school) RP, altered the association between student race and receipt of out-of-school suspensions. Students' participation in RP reduced the odds of receiving out-of-school suspensions, with similar benefits reported across racial groups.

Turning to students' perspectives of their teachers' implementation of RP, Gregory and colleagues (2015, 2016) found that Latino and African American students who rated their teachers with high RP implementation were issued fewer exclusionary discipline referrals compared with students who rated their teachers with low RP implementation. Levels of referrals for misconduct or defiance were lower during the RP implementation years when compared to the two preceding years (Gregory & Clawson, 2016).

### 3.5.2. Interpersonal student outcomes: bullying and aggression

Three studies focused on bullying in experimental designs. In a cluster RCT with grade six and seven students, Acosta and colleagues (2019) reported no differences regarding bullying between the experimental and the control group after implementing RP for two years. However, students' self-reported experience with RP was associated with reduced physical and cyberbullying. Another cluster RCT by Bonell et al. (2018) examined the effects of the Learning Together intervention, which included RP, on bullying and aggression compared to standard practice. In the intervention group, students' reports of bullying victimisation compared with schools continuing their standard practice were reduced, and the difference remained significant after 36 months. Still, there was no difference regarding students' self-reported perpetration of aggressive behaviours towards other students. Using a quasi-experimental pre-post design to compare a whole-school RP approach, a partial implementation of RP and no intervention, Wong and colleagues (2011) found that bullying decreased in grades seven to nine of both intervention groups, whereas bullying increased in the control group.

## 3.6. Secondary outcomes

Turning to the secondary outcomes reported in the included studies, we identified four different outcome categories concerning findings

related to (a) characteristics associated with the school (e.g. school climate, safety, connectedness and attendance), (b) teachers' perspectives of implementing RP, (c) (intra-) personal factors, and (d) interpersonal relationships between teachers and students, as well as amongst students.

### 3.6.1. School climate, safety, connectedness and attendance

Six studies examined school-related issues like school climate, safety, connectedness and attendance rates. Findings from the cluster RCT by Acosta and colleagues (2019) showed that RP did not improve student ratings of school connectedness or school climate. However, students' self-reported experience with RP, assessed by asking how often their teachers used specific methods of RP (e.g. affective statements), was significantly associated with improved school climate and school connectedness at post-test. A critical aspect of this study is that according to the survey data only a minority of students in intervention schools experienced RP to a great extent, while students in control schools experienced much more RP than would be expected (Acosta et al., 2019). Accordingly, Augustine and colleagues (2018) also conducted a cluster RCT and reported tendencies of lower classroom climate ratings of students in the intervention group than in the control group after implementing RP.

Turning to students' feelings of safety in school, Morrison (2002) used a one-group pre-post design and found that after the implementation of the Responsible Citizenship Program, which is grounded in a number of principles of RP, year five students reported increased feelings of safety in the school community. However, this study did not clarify the extent to which the programme included RP.

Norris (2019) evaluated the impact of RP on school engagement. She conducted a quasi-experimental design to compare a reactive-only model of RP, a traditional whole-school approach, and a proactive-only whole-school approach in three different schools. The reactive-only model of RP did not lead to improved student reported school engagement scores, but a comparison of schools with either a traditional whole-school approach or with a proactive-only approach did yield significant differences in school engagement, with an increase demonstrated in the proactive-only group.

Meanwhile, Wong et al. (2011) found that grade seven to nine students enrolled in a whole-school RP approach did not demonstrate an increased sense of belonging or school harmony in comparison to students who received a partial RP implementation or those who received no restorative interventions. Within-group differences for these partial RP and control group, however, showed significant decreases in both sense of belonging and school harmony. Contrarily, in the whole-school approach group there was a slight increase in these outcomes.

Turning to student attendance rates, Stinchcomb and colleagues (2006) did report an improvement following the implementation of RP in one elementary school, there was no information on two other schools in this respect. Accordingly, Augustine and colleagues (2018) did report improved attendance rates for elementary students and students in the individualised education programme in the intervention group compared to the control group, but not in other K-12 age groups.

Furthermore, the cluster RCT by Augustine and colleagues (2018) was the only study, which focused on academic outcomes and found that academic outcomes worsened in grade 6-8 in treatment schools, whereas PSAT scores for 10th grade students increased.

### 3.6.2. Teachers' perspectives of implementing RP

Moving from school or class-level outcomes to outcomes on the teacher level, we identified three studies with quantitative outcomes presenting teachers' perspectives on RP methods. Teachers in the RCT by Augustine and colleagues (2018) answered a survey that concerned teaching and learning conditions. Teachers in the intervention group rated conduct management, work safety, teacher leadership, school leadership, and overall teaching and learning conditions significantly higher than teachers in control schools.

Rainbolt et al. (2019) generated an online-survey for teachers after three years of implementing RP focusing on how they currently implement RP. Two-thirds of the respondents stated that they often or almost daily used RP. Furthermore, teachers rated affective statements as being most effective. Brown (2017) conducted a one-group post-test only-design and focused on the impact of implementing community-building circles on a listening culture. The survey responses indicated that teachers and support staff rated that there was a positive listening culture between various groups in school (e.g. administrators, parents, students) after the implementation of community-building circles.

### 3.6.3. (Intra) personal factors

Five studies focused on several (intra) personal factors of students, like happiness, shame management, self-esteem, psychological well-being and recidivism-rates. The cluster RCT by Bonell et al. (2018) found RP improved students' quality of life and psychological well-being, but reported no significant group differences regarding emotional problems, conduct problems and hyperactivity. Furthermore, they found lower odds in the intervention schools concerning substance use. By contrast, Norris's (2019) findings of implementing RP were mixed: reactive RP did not improve happiness and a traditional whole-school approach even reduced happiness, whereas a proactive-only RP approach did lead to an increase of happiness. However, baseline happiness scores in the traditional whole-school approach schools were higher than in proactive-only schools.

Turning to students' self-esteem and empathy, Wong et al. (2011) found that only whole-school RP implementation led to improvements, whilst partial or no implementation did not significantly affect students' self-esteem and empathy.

Using a sample of 1636 discharged students from a special school focused on troubled and at-risk youth that uses RP, McCold (2008) examined the relationship between recidivism rates and lengths of programme participation. Correlational analysis indicate that recidivism rates were significantly related to length of programme participation, and this relationship held after controlling for some risk factors such as age, gender, race, and prior offending.

Morrison (2002) examined year five students' shame management strategies after participating in the Responsible Citizenship Program, including RP to an unknown extent. After the implementation of the Responsible Citizenship Program, students reported increases in the use of adaptive shame management skills, and a decrease in the use of maladaptive shame management strategies.

### 3.6.4. Interpersonal relationships

Implementing RP as a conflict prevention and management strategy, may have a positive impact beyond direct measures of conflict intensity and frequency, as well as beyond influencing intrapersonal factors of teacher and students. Indeed, the focus of restorative approaches on managing relationships suggests positive outcomes could be expected on an interpersonal level of relationships between students and their teachers, or between students themselves (Hendry, 2009).

Four studies focused on relationships between students and teachers due to the implementation of RP, of which two explored teachers' perspectives and two explored students' perspectives. No studies explored both parties' perspectives in the same study. The cluster RCT by Augustine et al. (2018) found that a large part of teachers in the intervention schools reported moderate to great improvements regarding their relationships to students. Following this, Rainbolt et al. (2019) conducted a teacher survey on how they currently rate student-teacher relationships after three years of RP implementation in one high school, and found teachers rating student-teacher-relationships as being respectful.

Turning to students' perspectives, Gregory et al. (2015) reported that student-rated high RP implementing teachers were perceived as more respectful by their students compared to their lower RP implementing colleagues. Contrarily, the quasi-experimental study of

Wong et al. (2011) revealed that there were decreases in positive perception of students toward teachers in every group (non-, partial or whole-school RP implementation), but these differences were only significant in the non- and partial implementing RP group between pre- and post-test.

Meanwhile, three studies investigated the influence of RP on relationships between students in experimental designs and registered similar findings. The cluster RCT by Acosta and colleagues (2019) found after two years of implementing RP in schools, students did not report more positive peer relationships and better social skills than students in control schools. However, students' self-reported experience with RP was significantly associated with improved peer attachment, and social skills. The only other cluster RCT (Bonell et al., 2018) also found no impact of RP on peer problems. Accordingly, Wong et al. (2011) found a decrease in hurting others in the partial implementing RP group, whereas there were no differences between pre- and post-test in either the whole-school implementation or the control group. Meanwhile, caring behaviour reportedly increased in the partial implementation group, decreased in the control group, and remained stable in schools adopting a whole-school RP approach.

### 3.7. Quality appraisal

Given the wide range of study designs and outcome variables of the included studies, a quality appraisal can help identify which findings may be more robust or reliable. Using the EPHP Tool (National Collaborating Centre for Methods & Tools 2008) for quality appraisal of the included studies, no studies were rated as strong, two studies were rated as being at moderate risk of bias and most studies ( $n = 15$ ) were rated as methodologically weak (Table 4). Noticeable, all included cluster RCTs used a matching process based on baseline data, such as demographic, academic, and disciplinary data to randomise intervention and control schools. Most studies lacked control conditions, with the exception of the four cluster RCTs and two quasi-experimental designs. Furthermore, quantitative data from pre-post comparisons were sparse, as several studies included only a single post-intervention measurement ( $n = 3$ ) or were designed as cross-sectional studies ( $n = 4$ ). Moreover, some studies did not report whether they assessed the outcomes through record linkage. By contrast, reporting standards were mostly high regarding data collection methods. Overall, there are differences regarding the risk of bias in the studies with a high risk in most of the studies.

The systematic search did identify some peer-reviewed grey literature, which mainly reported on the positive effects of RP. We could not identify instances of selective reporting amongst the included studies.

## 4. Discussion

This systematic review aimed to identify and summarise quantitative studies about the effectiveness of RP in school settings. The search results highlight that this is a relatively new and emerging area of research, with almost half of the seventeen included studies being published in the last two years. Our systematic search also retrieved a protocol for an ongoing cluster RCT study by Green and colleagues (2019), which examines the influence of RP on expulsions, suspensions, truancy, bullying, grade-point average, sense of safety, and teacher support; but as no results were published yet it was not yet included in this review. Study designs and methodologies varied widely and included correlational studies, school case studies, quasi-experimental designs, and cluster RCTs. In addition, the number of measuring points varied between one and three measuring points. This variability in study design and methodology makes it difficult to directly compare the outcomes of the different studies and underscores the need for more high quality research to enable a meta-analytic review of the evidence base.

Our findings indicate that the evidence base for the effectiveness of RP in school settings is greatest for student suspensions and behavioural referrals. The findings were more inconsistent and less conclusive for

**Table 4**  
Quality appraisal of included studies ( $n = 17$ ) according to the EPHP Tool<sup>a</sup>.

| First author, year      | Selection Bias | Study design                         | Confounders | Blinding | Data collection methods | Withdrawals and drop-outs | Global rating |
|-------------------------|----------------|--------------------------------------|-------------|----------|-------------------------|---------------------------|---------------|
| Acosta et al., 2019     | -              | +(cluster RCT)                       | +           | -        | +                       | -                         | -             |
| Anyon et al., 2014      | +              | -(cross-sectional)                   | +           | -        | +                       | NA                        | -             |
| Anyon et al., 2016      | +              | -(cross-sectional)                   | +           | -        | +                       | NA                        | -             |
| Augustine et al., 2018  | -              | +(cluster RCT)                       | +           | -        | -                       | -                         | -             |
| Bonell et al., 2018     | +              | +(cluster RCT)                       | +           | +        | +                       | -                         | 0             |
| Brown, 2017             | -              | -(one group, one time survey)        | -           | -        | -                       | NA                        | -             |
| Gregory et al., 2015    | -              | -(one group, one time survey)        | +           | -        | +                       | NA                        | -             |
| Gregory & Clawson, 2016 | -              | -(one group, one time survey)        | 0           | -        | +                       | NA                        | -             |
| Gregory et al., 2018    | +              | -(cross-sectional)                   | +           | -        | +                       | NA                        | -             |
| Hashim et al., 2018     | +              | 0 (interrupted time series analysis) | +           | -        | +                       | NA                        | 0             |
| Mansfield et al., 2018  | -              | 0 (cohort, one group pre + post)     | -           | -        | +                       | NA                        | -             |
| McCold, 2008            | +              | -(cross-sectional)                   | +           | -        | +                       | NA                        | -             |
| Morrison, 2002          | -              | 0 (cohort, one group pre + post)     | -           | -        | +                       | -                         | -             |
| Norris, 2019            | -              | 0 (cohort analytic)                  | -           | -        | +                       | -                         | -             |
| Rainbolt et al., 2019   | -              | -(one group, one time survey)        | -           | -        | -                       | NA                        | -             |
| Stinchcomb et al., 2006 | -              | 0 (cohort, one group pre + post)     | -           | -        | +                       | NA                        | -             |
| Wong et al., 2011       | 0              | 0 (cohort analytic)                  | +           | -        | +                       | -                         | -             |

EPHP Tool: Effective Public Health Practice Project Quality Assessment Tool for Quantitative Studies (National Collaborating Centre for Methods & Tools 2008).

+ = strong, 0 = moderate, - = weak, NA = not applicable.

<sup>a</sup> Full wording of the quality tool.

other school-related issues, such as attendance rates, school connectedness, school climate, sense of belonging, and school harmony. The use of only proactive methods may positively influence students' happiness and school engagement in comparison to traditional whole-school approaches which involve both proactive and reactive methods. Furthermore, there is tentative evidence that RP may be associated with improved shame management, improved self-esteem, and reduced rates of revictimisation. Finally, RP may have a positive impact on relationships between teachers and students, as well as those between students, particularly with regards to bullying. However, only one study focused on academic outcomes and found mixed results.

#### 4.1. Strengths and limitations

Strengths to this systematic review are the rigour and methodology with which it was conducted, and adherence to the PRISMA reporting guidelines for systematic reviews. Double reviewing of full-lengths articles and quality appraisal strengthens the findings. The current review is the first systematic review collating and synthesising quantitative evidence of RP in school settings.

Limitations of the study relate to the search strategy. Hence, some relevant studies may not have been included. Furthermore, we have not made any restrictions regarding the study design. Therefore we included weaker study designs, such as correlational and one-group designs, which do not give any indications of causal inferences. However, in the absence of more independent group designs and controlled studies, these studies at least give an indication towards whether RP may have beneficial outcomes and whether it might be an intervention worth pursuing.

#### 4.2. Implications and future directions

While this systematic review suggests that interest in adopting RP to address conflicts in school is increasing, the current evidence base is limited and inconclusive at best. The findings with regard to methodological quality highlight a significant need for additional and higher-quality research on this topic. Sample selection, randomisation and selection of appropriate control conditions form one of the greater challenges for researchers. However, it is often organisationally impossible to offer different interventions to students of the same school, so that randomisation in clusters must be carried out at school level (Chenot, 2009). Higher grade classes that are taught by multiple teachers thereby face an additional challenge, in particular because it should be expected that

RP are implemented in a consistent way by all teachers, as it can be assumed that consistency and predictability are likely to affect the intervention's effectiveness. This is not easy to achieve, not in the least because many studies do not provide sufficiently detailed descriptions of how RP were implemented. Consequently, comparability of findings across studies is hampered, also because there does not appear to be a common definition of which methods are essential to RP (Daly, 2002; Reimer, 2011; Sellman et al., 2014). Hence, the methods of RP which were implemented in the studies varied widely, if they have been explicitly mentioned at all. Accordingly, prior findings indicate that the implementation of RP varies widely across schools (Jain et al., 2014; McCluskey et al., 2008). This makes it difficult to compare study samples and leads to an increase in study heterogeneity. Future research could investigate the application of single methods from the RP approach regarding their outcomes in practice. This would be useful, for example, to better interpret the results of Norris (2019), who found a light evidence for a benefit of proactive RP. Accordingly, a key priority is to describe the methodological procedure of future studies more precisely in order to be able to make more concrete statements about the effectiveness of RP in reducing conflicts in school settings.

Four of the included studies assessed implementation fidelity by teachers. Bonell and colleagues (2018) assessed fidelity for a range of interventions (e.g. convening of action groups, schools' delivery of modules) 24 and 36 months after the implementation started. Using students and teachers reports of experiences with particular RP elements, Gregory and colleagues (2015) examined the degree of RP implementation. It was unexpected, that correlations of student-reported and teacher-reported implementation were not significantly related to one another. Gregory and Clawson (2016) only used students' perspectives on the experiences with RP to assess the degree of RP implementation. In their survey, Rainbolt and colleagues (2019) also asked to what extent respondents used RP in their everyday school life, to gain an insight into the extent to which teachers actually use RP. Future research would be strengthened by using indicators that measure multiple and multilevel factors related to implementation fidelity to clarify their findings.

Overall generalisability of this reviews' findings is limited, because most of the research was conducted in the USA and single research teams contributed multiple studies to this review. Hence, studies from other research teams in other geographical regions and countries are necessary to establish the potential effectiveness of RP in other educational climates and cultures. So far, there are few cross-cultural studies on cultural differences, e.g. regarding suspension rates and school climate ratings. Nonetheless, countries differ in their school policies and ap-

proaches to handling problem behaviour, as can be seen in higher rates of school suspensions in the US compared to Australia (Hemphill, McMorris, Toubourou, Herrenkohl, Catalano & Mathers, 2007). Hence, it would be helpful to examine whether specific countries or cultures benefit more from RP than others, given their cultural background.

Meanwhile, multiple studies did not specify the time between the implementation of RP and outcome assessment and for almost half of the studies this duration was limited to a period of one to two years. With regards to the implementation of RP, this can be considered a rather short time span to expect intervention effects to be observed. Hence, some studies may not have captured the full effect of RP implementation on student and teacher outcomes. In this respect, studies with significantly longer time spans are needed.

Further gaps in the literature also warrant attention. Up to now, there is some contradictory evidence of the impact of RP on school climate perceived by students and teachers (Acosta et al., 2019; Augustine et al., 2018). With regard to the profound influence that positive school climate can have on the mental and physical health of students and teachers (Grayson & Alvarez, 2008; Thapa, Cohen, Guffey & Higgins-D'Alessandro, 2013), it is important to examine if RP could help to improve school climate. Furthermore, there is initial evidence from qualitative case studies that RP also have a positive impact on the well-being of teachers (Kaveney & Drewery, 2011). Nonetheless, there is a lack of experimental studies that focus on the impact of implementing RP on teachers, such as stress, and work satisfaction. Based on the findings that teacher well-being impacts student learning (Gray, Wilcox & Nordstokke, 2017), it is even more necessary to focus on teachers experiences associated with the implementation of RP. Moreover, teacher and student outcomes could also influence each other. Using structural equation modelling, Collie, Shapka, and Perry (2012) demonstrated that teachers' perceptions of students' behaviour significantly predicted teacher job satisfaction, teachers' sense of stress and teaching efficacy.

## 5. Conclusion

Taken together, RP might decrease student suspensions and behaviour referrals and potentially could improve student and teacher-relationships. The increased research activity in recent years and four recent cluster RCTs (Acosta et al., 2019; Augustine et al., 2018; Bonell et al., 2018; Green et al., 2019) indicate that there is an awareness of the need for proof of effectiveness of RP in school settings. However, as an intervention to reduce conflicts the current evidence remains limited and further research with robust methods is required.

## Declarations of Competing Interest

None.

## Acknowledgements

We would like to thank Charlotte Weber for her contribution as a reviewer in the study selection process and quality appraisal.

## Funding source

The present study was funded by a grant to Bettina Amrhein, Stefan Fries, and Leen Vereenoghe from the German Federal Ministry of Education and Research (support code: 01NV1738, support line: Qualification of educational professionals for inclusive education [Qualifizierung der pädagogischen Fachkräfte für inklusive Bildung]). The sole responsibility for the article's contents lies with the authors.

## References

Acosta, J., Chinman, M., Ebener, P., Malone, P. S., Phillips, A., & Wilks, A. (2019). Evaluation of a whole-school change intervention: Findings from a two-year cluster-randomized trial of the restorative practices intervention. *Journal of Youth And Adolescence*, 48(5), 876–890. [10.1007/s10964-019-01013-2](https://doi.org/10.1007/s10964-019-01013-2).

- Acosta, J. D., Chinman, M., Ebener, P., Phillips, A., Xenakis, L., & Malone, P. S. (2016). A cluster-randomized trial of restorative practices: An illustration to spur high-quality research and evaluation. *Journal of Educational and Psychological Consultation*, 26(4), 413–430. [10.1080/10474412.2016.1217488](https://doi.org/10.1080/10474412.2016.1217488).
- Anfara, V. A., J. r., Evans, K. R., & Lester, J. N. (2013). Restorative justice in education: What we know so far. *Middle School Journal*, 44(5), 57–63. [10.1080/00940771.2013.11461873](https://doi.org/10.1080/00940771.2013.11461873).
- Anyon, Y., Gregory, A., Stone, S., Farrar, J., Jenson, J. M., McQueen, J., et al. (2016). Restorative interventions and school discipline sanctions in a large urban school district. *American Educational Research Journal*, 53(6), 1663–1697. [10.3102/0002831216675719](https://doi.org/10.3102/0002831216675719).
- Anyon, Y., Jenson, J. M., Altschul, L., Farrar, J., McQueen, J., Greer, E., et al. (2014). The persistent effect of race and the promise of alternatives to suspension in school discipline outcomes. *Children and Youth Service Review*, 44, 379–386. [10.1016/j.childyouth.2014.06.025](https://doi.org/10.1016/j.childyouth.2014.06.025).
- Augustine, C. H., Engberg, J., Grimm, G. E., Lee, E., Wang, E. L., Christianson, K., et al. (2018). Can restorative practices improve school climate and curb suspensions? An evaluation of the impact of restorative practices in a mid-sized urban school district. *RAND Corporation*. [https://www.rand.org/pubs/research\\_reports/RR2840.html](https://www.rand.org/pubs/research_reports/RR2840.html). accessed 12 February 2020.
- Bonell, C., Allen, E., Warren, E., McGowan, J., Bevilacqua, L., Jamal, F., et al. (2018). Effects of the learning together intervention on bullying and aggression in English secondary schools (INCLUSIVE): A cluster randomised controlled trial. *The Lancet*, 392(10163), 2452–2464 [http://dx.doi.org/10.1016/S0140-6736\(18\)32280-3](http://dx.doi.org/10.1016/S0140-6736(18)32280-3).
- Brown, M. A. (2017). Being heard: How a listening culture supports the implementation of schoolwide restorative practices. *Restor. Justice*, 5(1), 53–69. [10.1080/20504721.2017.1294792](https://doi.org/10.1080/20504721.2017.1294792).
- Chenot, J. F. (2009). Cluster-randomisierte Studien: Eine wichtige Methode in der gemeinmedizinischen Forschung. *Z. für Evidenz. Fortbild. und qual. im Gesundheitswesen*, 103(7), 475–480 in German <https://doi.org/10.1016/j.zeqf.2009.07.004>.
- Collie, R., Shapka, J., & Perry, N. (2012). School climate and socioemotional learning: Predicting teacher stress, job satisfaction, and teaching efficacy. *Journal of Educational Psychology*, 104, 1189–1204 <http://dx.doi.org/10.1037/a0029356>.
- Costello, B., Wachtel, J., & Wachtel, T. (2009). *The restorative practices handbook: For teachers, disciplinarians and administrators*. Bethlehem, Pennsylvania: International Institute for Restorative Practices.
- Cremin, H. (2007). *Peer mediation: Citizenship and social inclusion revisited: Citizenship and social inclusion in action*. *Peer mediation*. Maidenhead, New York: Open University Press.
- Daly, K. (2002). Restorative justice: The real story. *Punishm. & soc.*, 4(1), 55–79. [10.1177/1462474022228464](https://doi.org/10.1177/1462474022228464).
- Fronius, T., Darling-Hammond, S., Persson, H., Guckenburger, S., Hurley, N., & Petrosino, A. (2019). Restorative justice in US schools: An updated research review. *WestEd*. <https://jprc.wested.org/wp-content/uploads/2019/03/Rd-Literature-Review-Update-2019.pdf>. accessed 25 November 2019.
- González, T. (2012). Keeping kids in schools: Restorative justice, punitive discipline, and the school to prison pipeline. *Journal of Law & Education*, 41(2), 281–335.
- Gray, C., Wilcox, G., & Nordstokke, D. (2017). Teacher mental health, school climate, inclusive education and student learning: A review. *Canadian Psychology*, 58(3), 203–210. [10.1037/cap0000117](https://doi.org/10.1037/cap0000117).
- Grayson, J. L., & Alvarez, H. K. (2008). School climate factors relating to teacher burnout: A mediator model. *Teaching and Technical Education*, 24, 1349–1363 <http://dx.doi.org/10.1016/j.tate.2007.06.005>.
- Green, A. E., Willging, C. E., Zamarin, K., Dehaiman, L. M., & Ruiloba, P. (2019). Cultivating healing by implementing restorative practices for youth: Protocol for a cluster randomized trial. *International Journal of Educational Research*, 93, 168–176. [10.1016/j.ijer.2018.11.005](https://doi.org/10.1016/j.ijer.2018.11.005).
- Gregory, A., & Clawson, K. (2016). The potential of restorative approaches to discipline for narrowing racial and gender disparities. In R. J. Skiba, K. Mediratta, & M. K. Rausch (Eds.), *Inequality in school discipline* (pp. 153–170). New York: Palgrave Macmillan.
- Gregory, A., Clawson, K., Davis, A., & Gerewitz, J. (2015). The promise of restorative practices to transform teacher-student relationships and achieve equity in school discipline. *Journal of Educational and Psychological Consultation*, 26(4), 325–353. [10.1080/10474412.2014.929950](https://doi.org/10.1080/10474412.2014.929950).
- Gregory, A., Huang, F. L., Anyon, Y., Greer, E., & Downing, B. (2018). An examination of restorative interventions and racial equity in out-of-school suspensions. *School Psychology Review*, 47(2), 167–182. [10.17105/SPR-2017-0073.V47-2](https://doi.org/10.17105/SPR-2017-0073.V47-2).
- Grossi, P. K., & dos Santos, A. M. (2012). Bullying in Brazilian Schools and Restorative Practices. *Canadian Journal of Education*, 35(1), 120–136.
- Hashim, A. K., Strunk, K. O., & Dhaliwal, T. K. (2018). Justice for all? Suspension bans and restorative justice programs in the Los Angeles Unified School District. *Peabody Journal of Education*, 93(2), 174–189. [10.1080/0161956X.2018.1435040](https://doi.org/10.1080/0161956X.2018.1435040).
- Hemphill, S. A., McMorris, B. J., Toubourou, J. W., Herrenkohl, T. I., Catalano, R. F., & Mathers, M. (2007). Rates of student-reported antisocial behavior, school suspensions, and arrests in Victoria, Australia and Washington State, United States. *Journal of School Health*, 77(6), 303–311. [10.1111/j.1746-1561.2007.00211.x](https://doi.org/10.1111/j.1746-1561.2007.00211.x).
- Hendry, R. (2009). *Building and restoring respectful relationships in schools: A guide to using restorative practice*. Abingdon: Routledge.
- Hopkins, B. (2004). *Just schools: A whole school approach to restorative justice*. London: Jessica Kingsley.
- International Institute for Restorative Practices, SaferSanerSchools™ whole-school change through restorative practices. <https://www.iirp.edu/pdf/WSC-Overview.pdf>, (2011). (assessed on 7 July 2018).
- Jain, S., Bassey, H., Brown, M. A., & Kalra, P. (2014). Restorative justice in Oakland schools: Implementation and impacts. *An Effective Strategy to Reduce Racially Disproportionate Discipline, Suspensions, and Improve Academic Out-*

- comes <https://www.ousd.org/cms/lib/CA01001176/Centricity/Domain/134/OUUSD-RJ%20Report%20revised%20Final.pdf> accessed 10 October 2019.
- Kane, J., Lloyd, G., McCluskey, G., Maguire, R., Riddell, S., Stead, J., et al. (2009). Generating an inclusive ethos? Exploring the impact of restorative practices in Scottish schools. *International Journal of Inclusive Education*, 13(3), 231–251. [10.1080/13603110701412950](https://doi.org/10.1080/13603110701412950).
- Kaveney, K., & Drewery, W. (2011). Classroom meetings as a restorative practice: A study of teachers' responses to an extended professional development innovation. *International Journal on School Disaffection*, 8(1), 5–12.
- Kline, D. M. S. (2016). Can restorative practices help to reduce disparities in school discipline data? A review of the literature. *Multicultural Perspective*, 18(2), 97–102. [10.1080/15210960.2016.1159099](https://doi.org/10.1080/15210960.2016.1159099).
- Lewis, S. (2009). Improving school climate: findings from schools implementing restorative practices. *International Institute for Restorative Practices*. <https://www.iirp.edu/pdf/IIRP-Improving-School-Climate-2009.pdf>. accessed 14 August 2018.
- Mansfield, K. C., Fowler, B., & Rainbolt, S. (2018). The potential of restorative practices to ameliorate discipline gaps: The story of one high school's leadership team. *Educational Administration Quarterly*, 54(2), 303–323. [10.1177/0013161X17751178](https://doi.org/10.1177/0013161X17751178).
- McCluskey, G., Lloyd, G., Kane, J., Riddell, S., Stead, J., & Weedon, E. (2008). Can restorative practices in schools make a difference? *Educational Review*, 60(4), 405–417. [10.1080/00131910802393456](https://doi.org/10.1080/00131910802393456).
- McCold, P. (2008). Evaluation of a restorative milieu: Restorative practices in context. In H. Ventura Miller (Ed.), *Restorative justice: From theory to practice* (pp. 99–137). Bingley, UK: Emerald Group Publishing Limited.
- McMorris, B.J., Beckman, K.J., Shea, G., Baumgartner, J., & Eggert, R.C. (2013). Applying restorative justice practices to Minneapolis public schools students recommended for possible expulsion: A pilot program evaluation of the family and youth restorative conference program, Final Report. [http://www.legalrightscenter.org/uploads/2/5/7/3/25735760/lrc\\_umn\\_report-final.pdf](http://www.legalrightscenter.org/uploads/2/5/7/3/25735760/lrc_umn_report-final.pdf), (accessed 6 January 2019).
- Mirsky, L. (2007). SaferSanerSchools: Transforming school cultures with restorative practices. *Reclaiming Child and Youth*, 16(2), 5–12.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. for the PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *British Medical Journal (online)*, 339(7716), 332–336. [10.1136/bmj.b2535](https://doi.org/10.1136/bmj.b2535).
- Morrison, B. (2002). Bullying and victimization in schools: A restorative justice approach, Trends & issues in crime and crim. *Justice*, 219, 1–6.
- Morrison, B. E., & Vaandering, D. (2012). Restorative justice: Pedagogy, praxis, and discipline. *Journal of School Violence*, 11(2), 138–155. [10.1080/15388220.2011.653322](https://doi.org/10.1080/15388220.2011.653322).
- National Collaborating Centre for Methods and Tools, Quality assessment tool for quantitative studies. [https://merst.ca/wp-content/uploads/2018/02/quality-assessment-tool\\_2010.pdf](https://merst.ca/wp-content/uploads/2018/02/quality-assessment-tool_2010.pdf), (2008). (accessed 24 April 2020).
- Norris, H. (2019). The impact of restorative approaches on well-being: An evaluation of happiness and engagement in schools. *Conflict Resolution Quarterly*, 36(3), 221–234. [10.1002/crq.21242](https://doi.org/10.1002/crq.21242).
- Rainbolt, S., Fowler, E. S., & Mansfield, K. C. (2019). High school teachers' perceptions of restorative discipline practices. *NASSP Bulletin*, 103(2), 158–182. [10.1177/0192636519853018](https://doi.org/10.1177/0192636519853018).
- Reimer, K. (2011). An exploration of the implementation of restorative justice in an Ontario public school. *Canadian Journal of Educational Administration and Policy*, 119, 1207–7798.
- Sellman, E., Cremin, H., & McCluskey, G. (2014). *Restorative approaches to conflict in schools*. London: Routledge.
- Stinchcomb, J. B., Bazemore, G., & Riestenberg, N. (2006). Beyond zero tolerance: Restoring justice in secondary schools. *Youth Violence and Juvenile Justice*, 4(2), 123–147. [10.1177/1541204006286287](https://doi.org/10.1177/1541204006286287).
- Sumner, M. D., Silverman, C. J., & Frampton, M. L. (2010). School-based restorative justice as an alternative to zero-tolerance policies: Lessons from West Oakland, Thelton E. Henderson Center for Social Justice. [https://www.law.berkeley.edu/files/thcsj/10-2010\\_School-based\\_Restorative\\_Justice\\_As\\_an\\_Alternative\\_to\\_Zero-Tolerance\\_Policies.pdf](https://www.law.berkeley.edu/files/thcsj/10-2010_School-based_Restorative_Justice_As_an_Alternative_to_Zero-Tolerance_Policies.pdf). accessed 10 October 2018.
- Thapa, A., Cohen, D., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, 83, 357–385 <http://dx.doi.org/10.3102/0034654313483907>.
- Wachtel, T. (2013). Defining restorative, international institute of restorative practices. <http://www.iirp.edu/pdf/Defining-Restorative.pdf>, (accessed 3 July 2018).
- Wong, D. S., Cheng, C. H., Ngan, R. M., & Ma, S. K. (2011). Program effectiveness of a restorative whole-school approach for tackling school bullying in Hong Kong. *International Journal of Offender Therapy and Comparative Criminology*, 55(6), 846–862. [10.1177/0306624X10374638](https://doi.org/10.1177/0306624X10374638).
- Zehr, H. (2002). *The little book of restorative justice*. Intercourse, PA: Good Books.