Common Practice Elements in Treatment Programs for Adolescents with Externalizing and Internalizing Problems: A Meta-Analysis


To link to this article: https://doi.org/10.1080/0886571X.2022.2111397

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.
Common Practice Elements in Treatment Programs for Adolescents with Externalizing and Internalizing Problems: A Meta-Analysis

Line Solheim Kvamme¹, Serap Keles², Ragnhild Bang Nes³,⁴, Linda Vaskinn⁵, Pamela M. Waaler⁶,⁷, Tore Wentzel-Larsen⁸,⁹, and John Kjøbli⁹,¹⁰

¹Center for Child and Adolescent Mental Health, Oslo, Norway; ²Knowledge Centre for Education, University of Stavanger, Stavanger, Norway; ³PROMENTA Research Center, Department of Psychology, University of Oslo, Oslo, Norway; ⁴Department of Mental Health and Suicide, Norwegian Institute for Public Health, Oslo, Norway; ⁵The Norwegian Directorate for Children, Youth and Family Affairs, Oslo, Norway; ⁶Department of Special Needs Education, University of Oslo, Oslo, Norway; ⁷Department of Education, University of Oslo, Oslo, Norway

ABSTRACT

There is a pressing need for effective interventions targeting mental disorders for service-involved youth across child welfare, correctional and mental health residential settings. Identifying effective common practice elements (CPEs) is a promising direction toward developing effective, flexible, and feasible therapeutic interventions. The aims of this study were: 1) to identify CPEs in treatment programs for adolescents with internalizing and/or externalizing disorders across residential settings, 2) to identify which CPEs are present in effective versus all trials, and 3) to estimate whether the presence of each CPE is associated with treatment effects. A systematic search identified 24 trials of programs targeting adolescent internalizing and/or externalizing symptoms, yielding 148 effect estimates. Discrete practices were coded into CPE categories across effective trials and the full sample. Eleven CPEs were identified and employed as moderators in three-level meta-analyses. We found large significant moderator effects of two elements on externalizing symptoms: Emotion recognition and differentiation, and Psychoeducation. No CPE was significantly related to effects on internalizing outcomes. The results illustrate the difference between CPE frequency and association with effect, favoring a combination of the two approaches. Emotion recognition and differentiation, and Psychoeducation should be prioritized in novel interventions targeting adolescent externalizing behavior in residential settings.

KEYWORDS

Mental health; Meta-Analysis; adolescents; common elements

Background

Dozens of empirically supported programs have been developed to prevent and reduce adolescent mental health problems (Weisz & Kazdin, 2017; Weisz et al., 2017). For adolescents who receive treatment in mental health, child
welfare, or forensic residential settings (hereafter referred to as “residential settings”), co-occurring emotional and behavioral disorders are the rule rather than the exception (Jozeﬁak et al., 2016; Leloux-Opmeer et al., 2016; Souverein et al., 2019). Across residential settings, adolescents have high rates of similar and co-occurring disorders; in residential group care (Jozeﬁak et al., 2016), foster care (Vaughn et al., 2008), mental health care (Deas, 2006), and forensic youth care (Beaudry et al., 2021). Adolescents with co-occurring disorders may be placed in, or voluntarily be admitted to, a range of settings in which they receive treatment of varying intensity, quality, and duration. The defining features of these facilities are shaped by the structure and jurisdictions of local and/or national welfare, health, and juvenile justice agencies. Several efforts have been made to classify residential treatment settings and to define the terms used in this field of research (Evenboer et al., 2012; Lee & Barth, 2011; Whittaker et al., 2016). These efforts may have moved the field toward consensus on reporting standards and definitions but have yet to permeate it fully. The differences in “who goes where” across states and systems (Farmer et al., 2016), and the limited consensus on definitions have made it challenging to conduct comparative studies of the efficacy of treatment programs for service-involved youth (Janssen-de Ruijter et al., 2017; Kvamme et al., 2021; Van Breda et al., 2021; Young et al., 2017).

Another challenge, given the high comorbidity rates of this population, is that many empirically supported programs target a single mental health problem. In a meta-analysis of 447 trials investigating the effects of psychotherapy for youth (albeit 303 of which were conducted in North America), only ten addressed multiple mental health problems (Weisz et al., 2017). Research efforts to determine the efficacy of interventions for service-involved adolescents have primarily focused on family and community-based interventions, which, according to some, may have slowed the development of effective treatment options for youth in residential settings (Whittaker et al., 2016). However, scholars have voiced the need for high-quality therapeutic interventions in residential settings, including forensic youth care, as part of the spectrum of treatment options for adolescents with severe and complex problems (Boel-Studt & Tobia, 2016; Daly et al., 2018; Farmer et al., 2017; James et al., 2006; Whittaker et al., 2016). Some have argued that the dichotomy of family-based and residential care is unhelpful, and that the focus instead should be on how to tailor interventions (Bellonci & Holmes, 2021). Tailored interventions must be adapted to the specific residential setting (e.g., Lewis & Marriott, 2018), and could be applied to a wide range of populations, including recipients of mental health and forensic youth care.

According to a meta-analysis including samples from 19 nations, the prevalence of psychotic disorders, major depression, PTSD, and ADHD are more than double among adolescents in forensic institutions, compared to their non-detained peers, and they are seven times more likely to have a conduct disorder
(Beaudry et al., 2021). It has been estimated that 80% of U.S. adolescents in foster care have behavioral problems, and 90% of those with externalizing disorders are arrested within one year of leaving care (Vaughn et al., 2008). Similarly, 75% of U.S. adolescents in forensic youth care are arrested within three years of reentering their communities (Mendel, 2015). In Swiss forensic youth care, psychotherapeutic interventions are widely used, regardless of whether placements are instigated by civil or criminal justice agencies (Souverein et al., 2019). In the Netherlands, forensic health care institutions offer treatment for adolescents due to the high rates of co-occurring mental health and conduct problems among young offenders (Colins et al., 2010; Hillege et al., 2018). The composition of co-occurring disorders in adolescents across residential settings is strikingly similar. These youth likely comprise a significant proportion of the so-called 80/20 rule of social cost, denoting that 20% of the population represent 80% of behavioral and health problems (Caspi et al., 2017). The WHO has called the potential of successful health interventions during adolescence a “triple dividend,” as it can improve the young person’s current health and wellbeing, creating a better foundation for their adult life, which in turn will “trickle down” on their future children (World Health Organization, 2022). The highest possible potential for a triple dividend pay-off arguably lies in effective interventions for adolescents in residential settings, given their high rates of comorbidity and adverse life outcomes. Complex and co-occurring disorders call for tailored, flexible treatment options. This could also be a possible route toward overcoming disparities within and between child welfare, health care, and juvenile justice facilities across nations and regions. Taken together, this points toward possible gains of investigating which practices are common across treatment programs in residential settings, and which practices are associated with beneficial outcomes.

One approach to unraveling mechanisms of change in treatment programs, is to identify discrete techniques, or common elements, used across therapeutic interventions (e.g., Chorpita et al., 2005; Leijten et al., 2019; Mulder & Rucklidge, 2017). Some common elements used in reduction and prevention of adolescent mental health problems have already been identified (Boustani et al., 2015; Chorpita & Daleiden, 2009; Rith-Najarian et al., 2019). The transdiagnostic and flexible nature of common elements is of major relevance for treatment of adolescent mental health and conduct problems across residential settings. Practices, mandates, and treatment intensity varies greatly across residential settings and systems, but the adolescents they serve share characteristically high rates of co-occurring problems. Therefore, we believe that identifying common practices, and whether they are associated with beneficial outcomes across forensic, mental health, and child welfare residential facilities could be a promising avenue for further research and development of tailored therapeutic interventions along the spectrum of residential settings.
Identifying Common Practice Elements Associated with Beneficial Effects

The term “common practice element” (CPE) was coined by Chorpita et al. (2005), defined as a “discrete clinical technique or strategy (e.g., “time out,” “relaxation”) used as part of a larger intervention plan” (p. 11). Treatment programs consist of multiple practice elements that should be associated with the overall effect of the program. One approach to identify CPEs across studies is the distillation and matching model (DMM; Chorpita & Daleiden, 2009; Chorpita et al., 2005). This entails systematic coding of CPEs across interventions that have been reported as effective for a certain treatment population and/or setting. Although the number of publications identifying CPEs has increased significantly in the last decade, knowledge is scarce about which elements are essential, which elements are ineffective, and which elements are potentially counterproductive (Leijten et al., 2021). This is vital information when developing novel and more optimized interventions. Some studies have examined the effects of CPEs in psychosocial interventions, but to the best of our knowledge none have addressed programs for adolescents in residential settings. The lack of knowledge about the effectiveness of program elements hampers improvement of interventions and makes it difficult to build on what others have done in previous trials (Leijten et al., 2021).

Building on Previous Element Approaches and Moving Forward

The DMM approach identifies CPEs across effective treatments, providing information about the presence and frequency of CPEs (e.g., Boustani et al., 2020; Rith-Najarian et al., 2019). Meta-analyses provide information about the overall effect of interventions, but they do not distinguish between program elements unless the elements are coded and included as moderators or predictors in the analyses. Identifying CPEs that may contribute to treatment effect is, de facto, defining the independent variables (Chorpita & Daleiden, 2009). Therefore, meta-analyses are a viable methodological option for making predictions about which CPEs contribute to treatment effects. In recent years, several reviews of elements have been conducted for several types of child and adolescent problems, employing increasingly advanced analytical approaches, including multilevel meta-analyses (see for instance, Boustani et al., 2020; Brown et al., 2017; Engell et al., 2020; Hogue et al., 2019, 2019; Leijten et al., 2019).

In a recent scoping review, Leijten et al. (2021) described and assessed various approaches to identifying common effective elements. In it, the authors reaffirm that reporting elements across trials with large overall effect estimates can only provide limited information for explaining treatment effects. If the CPE approach is to unpack the black box of treatment programs, information about how each CPE is associated with treatment effects is necessary (Mulder & Rucklidge, 2017). Multilevel meta-, moderator-, and
subgroup analyses are promising tools for identifying how the presence or absence of individual elements predict changes in the overall effect of a treatment program (Leijten et al., 2021). These are also suitable approaches to identify transdiagnostic elements (i.e., elements associated with positive outcomes across disorders), and to develop hypotheses to be tested in novel therapeutic interventions (Marchette & Weisz, 2017).

As noted, we are not aware of any extant studies of CPEs in programs for adolescents in residential settings. Using a CPE approach has been proposed as one promising and feasible way toward better treatment for this population (Lee & McMillen, 2017). Given the limited amount of high-quality randomized controlled trials (RCTs) on this topic, deriving CPEs from studies that do exist may be a fruitful way to utilize these data. The present study applies two meta-analytic approaches to identifying CPEs, one of which reveals which elements are common for and present in effective treatment programs (inspired by DMM). The other provides information on whether each CPE is associated with treatment effects across all studies, including those reporting iatrogenic or null effects. The latter aligns with what has been proposed as a necessary paradigm shift from branded programs toward tailored interventions (Hofmann & Hayes, 2019).

The Present Study

The aim of this study is to identify CPEs in therapeutic interventions for youth in residential settings by examining: 1) which practice elements are present across treatment programs targeting externalizing or internalizing symptoms in residential settings (i.e., which practice elements are common); 2) which CPEs are present across effective treatment programs; and 3) whether any CPEs predict beneficial outcomes on internalizing or externalizing problems, or both, across programs and residential settings.

Method

Literature Search

We used a search strategy to be used in a systematic review. The search was conducted in 2019 and updated in April 2021. The systematic review is under preparation and will be referred to as Study 1 hereafter. The study protocol for Study 1 has been pre-registered in the International Prospective Register of Systematic Reviews (Prospero) and is available under the record ID CRD42019126853. The pre-registered protocol for the present study is also available from Prospero, using the record ID CRD42021245319. Publications from the following bibliographic databases were obtained: MEDLINE (Ovid), PsycINFO (Ovid), Cochrane Library Central, Sociological, Abstracts...
Inclusion and Exclusion Criteria

We identified RCTs of treatment programs for adolescents between the age of 12–18 with internalizing and/or externalizing symptoms in residential settings. Studies with participants who had chronic somatic diseases or developmental disorders were excluded. Studies that did not report effect sizes pertaining to internalizing and/or externalizing symptoms were excluded. Residential settings across youth-serving systems, i.e., child welfare, juvenile justice, mental health, were eligible for inclusion. These disparate treatment settings cater to many of the same (co-occurring) problems. Therefore, we deemed it relevant to identify CPEs across treatment settings, and their associations with treatment effects. Studies of educational or somatic settings, and orphanages were excluded. No exclusion criteria pertaining to publication year, type, or language were employed.

Data Extraction and Coding

In Study 1, information about study sample (e.g., gender composition), type of informant (e.g., youth self-report, facility staff, teacher reports), effect sizes (ES), intervention and design characteristics were extracted. We obtained the post-measurement sample sizes for all studies, except for one study (Schlichter & Horan, 1981), in which 27 participants were randomized to two treatment conditions and TAU. We used the randomized sample sizes for this study, as post-measure n or attrition was not systematically reported. In addition, we extracted the following data: target outcomes (i.e., internalizing, or externalizing symptoms), measurement characteristics (i.e., clinical questionnaires or behavior measurements reported), control group characteristics (i.e., active, or passive control group(s)), and the number of treatment arms reported for each study. Then we coded the discrete practices employed in each study and each treatment arm, as described in the study program manual, article, and/or supplementary materials. We developed a codebook based on a template used to identify CPEs in emotion regulation interventions for children and adolescents (Espenes et al., 2022; Helland et al., 2022).

Two doctoral level psychology researchers and four master’s level research group members revised the codebook. First, a selection (N = 20) of descriptions of therapeutic practices was examined independently by the coders. Each practice was assigned to existing or new codes, until no further codes were suggested. Second, the codes were grouped to comprise “common practice
elements” (CPEs). The elements represent the breadth of practices that could be applied within a category (e.g., the element Mindfulness, Awareness and Relaxation comprise practices such as Breathing exercise or Awareness of bodily sensations). See Supplementary Table 1 for a detailed description of codes and elements. Third, the codebook was applied to the remaining studies by indicating whether a practice was present or absent in each study.

Elements were coded as “present” or “not present” for each study. For trials with several comparison groups, we coded elements as present for each treatment arm the element was used in. We developed a protocol defining a CPE as any element that was coded as present at least twice for each of the outcome categories Internalizing and Externalizing. Coding of each study was performed individually and then in dyads (coders were six master- or doctoral level psychology research team members). Discrepancies in coding were resolved by discussion and/or by consulting a third team member. Interrater reliability among coders were calculated for the elements that were eligible for final inclusion. The Fleiss’ kappas ranged between .43 and .93, which is considered moderate to almost perfect (Fleiss, 1981).

**Data Synthesis and Analyses**

The sample sizes, means and standard deviations reported in the included studies were converted into Cohen’s $d$ values and associated variance measures (squared standard errors). T-test statistics or $p$-values and post-measurement sample sizes were used to calculate Cohen’s $d$ when means and standard deviations were not reported. Cohen’s $d$ variance measures were calculated in accordance with Formula Three, as described in Marfo and Okyere (2019). We included multiple ESs per study for those that included multiple measures of internalizing or externalizing behavior, or both. Effect sizes indicating beneficial effects by reduction of the estimate number were converted so that all ESs were aligned in that a greater number reflected more beneficial effects.

To answer the research questions, two concurrent analytic strategies were employed. First, we identified “winning” treatment groups. Winning groups were defined as programs or comparison groups that either yielded significant overall treatment effects, or significantly outperformed at least one comparison group at the $p < .05$ level (Brown et al., 2017; Chorpita & Daleiden, 2009). Next, an application in the R package shiny (Chang et al., 2021) was used for the meta-analysis: for estimating overall effects of CPEs, heterogeneity, and moderator analyses. The application was constructed by author TWL, based on the Three-Level Meta-Analytic Model as presented by Assink and Wibbelink (2016). R code for the shiny app is available in the GitHub repository https://github.com/ToreWentzel-Larsen/threelevel.
First, an overall ES was estimated separately for each outcome (i.e., internalizing, or externalizing symptoms). The Cohen’s $d$ values were based on post measurements for each outcome in individual studies (i.e., ES between trials that compare a program against a control after the intervention period ended). A positive ES indicated that the intervention group fared better than the control group on the outcome variable. Some studies provided multiple ESs (e.g., multiple treatment or control groups, or effects on multiple measures of outcome variables). Therefore, we employed a three-level random-effects modeling approach that accounts for the dependency (nesting) of ESs in studies (Assink & Wibbelink, 2016; Cheung, 2014). Second, we tested whether the presence of the CPEs were associated with program effects by performing a three-level mixed effects meta-analysis. Each CPE was included as a moderator, which might also have contributed to the heterogeneity of ES (Cheung, 2014). The difference in ES between trials of therapeutic programs with the CPE versus trials of programs without the CPE was estimated. In addition, and in line with the protocol, the moderators pertaining to demographic makeup of the studies were examined. The moderators were gender composition of study participants, type of informant (youth, residential staff, caregiver, or teacher), and type of control group (active or passive).

The Q statistic was estimated to assess the heterogeneity of the ES across studies. A significant Q value rejects the assumption of homogeneity and shows the heterogeneity across studies. In addition, the $I^2$ statistic was computed to examine the proportion of variation in the ES that reflected true variation rather than sampling error (Borenstein et al., 2017). Potential publication bias in a three-level model on ES results was assessed by examining the funnel plot and running the modified Egger’s Regression Test (Egger et al., 1997; Marengo & Montag, 2020).

Results

Characteristics of Included Studies

Based on the search strategy employed for Study 1, 24 studies (total N = 1854, 148 ESs) matched our search criteria. The coding and classification of outcomes yielded fourteen studies and 38 ESs for internalizing outcomes and nineteen studies and 110 ESs for externalizing outcomes. Eight studies provided ESs for both outcome categories. Twenty studies (83%) were conducted in the USA, the remaining four (17%) in Canada, England, Portugal, and Germany. Seven studies (29%) were conducted in mental health care facilities, one study (4%) in a child welfare setting, 13 studies (54%) in juvenile justice/correctional facilities, and three studies (12%) reported on facilities that were a mix of child welfare and mental health services (see, Table 1 for details). Participants’ mean age was 15 years (SD = 1.2), 13 studies (52%) included boys
and girls, 10 studies (42%) included boys only, and one study included girls only. Publication year ranged from 1981 (Schlichter & Horan, 1981) to 2020 (Hein et al., 2020; Pereira et al., 2020), the average number of years since the studies were published was 17. Supplementary Table 2 displays the excluded studies and the reasons for exclusion, and status for solicitation of additional information from study authors. The most common reason for exclusion was insufficient reports of information to calculate ES (N = 24). Three studies (Greenbaum & Javdani, 2017; Talley, 2013; and Wang, 2021) were excluded due to the reported measurements being incompatible with the outcome categories internalizing and/or externalizing behavior. Table 1 displays all primary studies, interventions, residential settings, and reported outcome measurements.

**Common Practice Elements**

Eleven CPEs were identified across studies, all made up of discrete practices as described in the primary studies. Supplementary Table 1 displays the CPEs, the discrete practices they comprised, and the number of times they were coded as present across studies, treatment arms, and outcome measures (given that most studies reported several measures of internalizing and/or externalizing outcomes, for each treatment arm). The most common elements were problem solving skills (n = 42), preventing maladaptive responses to emotional distress (n = 32); and social skills training (n = 31). The element psychoeducation was only coded once for internalizing outcomes and thus deviated from the predefined cutoff (i.e., coded as “present” twice or more for both outcome categories). It was retained for analyses regarding externalizing outcomes only (present in eight studies, 18 ESs). Figure 1 and Figure 2 display the number of times each element was coded as present across studies for externalizing and internalizing outcomes, respectively. The figures differentiate between element frequencies for the total sample (24 studies, 148 ES) and that of winning groups (N = 19 studies, 84 ESs). The p-values indicate that there were significant differences in the presence of a CPE for winning groups compared to all included ES.

**Element Frequencies across Winning and Non-Winning Groups**

There were significant differences in the presence of elements when comparing winning and non-winning interventions. Supplementary Table 4 presents results of analyses of variance (ANOVA) for each element, grouped for outcome, across winning and non-winning interventions. For externalizing outcomes, large significant differences between winning and non-winning interventions were found for the elements training in behavior regulation, and psychoeducation. Moderate to large differences were found for preventing
Table 1. Included primary studies, treatment setting, and outcome measurements.

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Intervention</th>
<th>Treatment setting</th>
<th>Duration: weeks (sessions/week)</th>
<th>Internalizing measurement</th>
<th>Externalizing measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahrens and Rexford (2002)</td>
<td>Cognitive Processing Therapy (CPT)</td>
<td>Juvenile Justice/corrections (“youth facility for adolescent offenders”)</td>
<td>8 (1)</td>
<td>PTSD Symptom Scale Self-Report (PSS-SR); Beck Depression Inventory (BDI)</td>
<td>N/A</td>
</tr>
<tr>
<td>Bunford (2017)*</td>
<td>Interpersonal Skills Group – Corrections Modified (ISG-CM)</td>
<td>Juvenile Justice/corrections (Perry Multi-County Juvenile Facility for “detained juvenile offenders”)</td>
<td>7 (1.7)</td>
<td>Difficulties in Emotion Regulation (DEIS)</td>
<td>Disruptive Behavior Disorders Rating Scale-Parent report (DBD); Modified Overt Aggression Scale (MOAS); Disruptive Behavior Disorders Rating Scale DBD-Teacher</td>
</tr>
<tr>
<td>Cianciolo (1989)*</td>
<td>Skills Training groups (problem solving, cognitive reappraisal, behavioral rehearsal, and relaxation) to reduce “acting out behavior”</td>
<td>Juvenile Justice/corrections (New York State youth facility, adolescents “placed there by the court system”)</td>
<td>6 (2)</td>
<td>State Trait Anxiety Inventory</td>
<td>Behavioral Measures: Incident Reports</td>
</tr>
<tr>
<td>Cohen-Kahn (1995)</td>
<td>“Muscle Max” tailored weight-training program, using graded mastery to reduce depression</td>
<td>Mental health (“residential treatment center” for adolescent inpatients)</td>
<td>8 (3)</td>
<td>Beck Depression Inventory (BDI); Multiple Affect Adjective Check List (MAACL-R) (depression)</td>
<td>N/A</td>
</tr>
<tr>
<td>Coleman et al. (1992)</td>
<td>Aggression Replacement Training (ART)</td>
<td>Mental health/Child welfare (“residential treatment center” for adolescents with behavioral disorders. 54% of study sample were under state custody)</td>
<td>10 (5)</td>
<td>N/A</td>
<td>The Behavior Incident Report (BIR); Self-Control Rating Scale (SCRS)</td>
</tr>
<tr>
<td>Ellmann (2002)*</td>
<td>Problem Solving, Cognitive Restructuring, and Combined Problem Solving/Cognitive Restructuring</td>
<td>Mental health (“residential treatment center” for adolescents with conduct disorders and ADHD)</td>
<td>4 (2)</td>
<td>N/A</td>
<td>The State-Trait Anger Expression Inventory (STAXI); Anger Expression Subscale; Child Behavior Checklist; Belief in Hostility/Aggression test</td>
</tr>
<tr>
<td>Gest et al. (2015)</td>
<td>Combined wake and light therapy (COWALI) to reduce depressive symptoms</td>
<td>Mental health (psychiatric hospital for adolescent psychiatric inpatients)</td>
<td>2 (5)</td>
<td>Beck Depression Inventory-II (BDI-II); The Strength and Difficulties Questionnaire (SDQ);</td>
<td>N/A</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Intervention</th>
<th>Treatment setting</th>
<th>Duration: weeks (sessions/week)</th>
<th>Internalizing measurement</th>
<th>Externalizing measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldstein et al. (2018) *a</td>
<td>Juvenile Justice Anger Management (JJAM)</td>
<td>Juvenile Justice/corrections (&quot;three residential juvenile justice facilities&quot;)</td>
<td>8 (2)</td>
<td>N/A</td>
<td>The Novaco Anger Scale and Orovation Inventory (NAS-PI); Aggression Questionnaire (AQ); The Intent Attributions and Feelings of Distress Measure (IAFD, Hostile Attributions subscale)</td>
</tr>
<tr>
<td>Hawkins et al. (1991) *a</td>
<td>Intensive Skills Training Program (skills training, social network development &amp; involvement in prosocial activities)</td>
<td>Juvenile Justice/corrections (a Washington state juvenile correctional facility housing adjudicated delinquent adolescents)</td>
<td>10 (2)</td>
<td>N/A</td>
<td>Adolescent Problem Situation Inventory (APSI); Drug and Alcohol Avoidance, social and problem solving, and self-control subscales</td>
</tr>
<tr>
<td>Hein et al. (2020)</td>
<td>Social problem-solving training (SPST)</td>
<td>Juvenile Justice/corrections</td>
<td>10 (1)</td>
<td>N/A</td>
<td>The State-Trait Anger Expression Inventory-2nd edition (STAXI); The Delinquency Reduction Outcome Profile (DROP)</td>
</tr>
<tr>
<td>Jackson (1991)*a</td>
<td>Aggression Replacement Training (ART)</td>
<td>Mental health (private psychiatric hospital, adolescents admitted for 28-day intensive treatment)</td>
<td>2 (5)</td>
<td>N/A</td>
<td>The Aggressive Behavior Checklist (ABC)</td>
</tr>
<tr>
<td>Jarden (1995)*a</td>
<td>Problem-Solving Group Therapy (PSGT) &amp; Problem-solving group therapy with generalization procedures (PSGT-GP)</td>
<td>Mental health/Child welfare (&quot;comprehensive treatment center offering residential services&quot; for adolescents with disruptive behavior disorder, referred by state departments of human services)</td>
<td>18 (1)</td>
<td>Child Behavior Checklist (CBCL); internalizing problems subscales</td>
<td>Child Behavior Checklist (CBCL); hyperactive, aggression, and delinquent subscales</td>
</tr>
<tr>
<td>Jewell and Elliff (2013) *a</td>
<td>Relaxation Skills Violence Prevention (RSVP)</td>
<td>Juvenile Justice/corrections (&quot;juvenile justice detention center&quot;)</td>
<td>1 (5)</td>
<td>N/A</td>
<td>Anger and Coping Questionnaire (ACQ)</td>
</tr>
<tr>
<td>Klatt (2008)*</td>
<td>Forgiveness intervention addressing previous victimization to reduce aggressive behavior through the healing of past wounds.</td>
<td>Juvenile justice/corrections (&quot;incarcerated youth at a juvenile treatment facility&quot;)</td>
<td>12 (1)</td>
<td>N/A</td>
<td>Aggression Questionnaire (AQ); Behavior observations</td>
</tr>
<tr>
<td>Author (year)</td>
<td>Intervention</td>
<td>Treatment setting</td>
<td>Duration: weeks (sessions/week)</td>
<td>Internalizing measurement</td>
<td>Externalizing measurement</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Ladley (1997)</td>
<td>Attributional-Intention Training to reduce anger and aggression</td>
<td>Mental health/Child welfare (&quot;Residential treatment program&quot; with average duration of ½-1 year, for adolescents “in placement” with socio-emotional difficulties and behavioral problems)</td>
<td>5 (2)</td>
<td>N/A</td>
<td>Teacher Checklist (TC): aggression subscale; Self-Control Rating Scale (SCRS)</td>
</tr>
<tr>
<td>Macmahon And Gross (1988)</td>
<td>Aerobic exercise to reduce depression</td>
<td>Juvenile Justice/corrections (&quot;incarcerated adolescents in juvenile detention facilities&quot;)</td>
<td>12 (3)</td>
<td>Beck Depression Inventory (BDI)</td>
<td>N/A</td>
</tr>
<tr>
<td>Pereira et al. (2020)</td>
<td>“Wave by wave” surf program with resilience-promoting components targeting self-esteem, self-regulation, healthy habits, and prosocial skills</td>
<td>Child welfare (adolescents living in residential childcare institutions)</td>
<td>21 (1)</td>
<td>Strengths and Difficulties Questionnaire (SDQ); Revised Children’s Anxiety and Depression Scale (RCADS)</td>
<td>N/A</td>
</tr>
<tr>
<td>Riemann et al. (2013)</td>
<td>Computerized Attention Modification (AMP), directing attention away from threatening stimuli</td>
<td>Mental health (&quot;residential treatment facility specializing in CBT for severe and complicated cases of anxiety disorders&quot;)</td>
<td>Not reported (7)</td>
<td>Screen for Child Anxiety Related Disorders (SCARED); Beck Depression Inventory II (BDI-II)</td>
<td>N/A</td>
</tr>
<tr>
<td>Rohde et al. (2004)</td>
<td>“Coping Course” Group-based CBT to enhance general coping and problem-solving skills</td>
<td>Juvenile Justice/corrections (&quot;Youth Correctional Facility providing closed custody care to male juvenile offenders&quot;)</td>
<td>8 (2)</td>
<td>Beck Depression Inventory II (BDI-II); Youth Self-Report (YSR): Internalizing</td>
<td>Youth Self-Report (YSR): Externalizing</td>
</tr>
<tr>
<td>Schlichter and Horan (1981)</td>
<td>Stress Inoculation through role-played provocations to reduce anger and aggression</td>
<td>Juvenile Justice/corrections (&quot;Adjudicated delinquents committed to a correctional facility&quot;)</td>
<td>5 (2)</td>
<td>N/A</td>
<td>Institutional behavior ratings: Physical and verbal aggression</td>
</tr>
<tr>
<td>Author et al. (year)</td>
<td>Intervention</td>
<td>Treatment setting</td>
<td>Duration: weeks (sessions/week)</td>
<td>Internalizing measurement</td>
<td>Externalizing measurement</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>--------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Schuurmans et al. (2018) *</td>
<td>“Dojo” emotion-evoking videogame with instructions in CBT-based relaxation techniques, positive thinking, and guided imagery</td>
<td>Mental health (Conducted in residential institutions that provide “secure youth care”)</td>
<td>4 (2)</td>
<td>Spence Children’s Anxiety Scale (SCAS): Child</td>
<td>Strengths and Difficulties Questionnaire (SDQ); externalizing subscales 'conduct problems,' 'hyperactivity-inattention,' and 'peer-problems.'</td>
</tr>
<tr>
<td>Stein et al. (2011)</td>
<td>Motivational interviewing (MI) to reduce depression and alcohol and marijuana consumption</td>
<td>Juvenile Justice/corrections (“state juvenile correctional facility”)</td>
<td>8 (2)</td>
<td>N/A</td>
<td>The risks and Consequences Questionnaire (RCQ)</td>
</tr>
</tbody>
</table>

Note. N/A = Not applicable *Dissertation *Study contained common practice elements significantly associated with increase in effect size.
Figure 1. Common practice element frequencies for externalizing outcomes: winning groups and all trials. Note: The bars represent the number of CPEs coded as “present” across comparisons. N = Total number of effect estimates for each comparison group. * p < .05 ** p < .01 *** p < .001

Figure 2. Common practice element frequencies for internalizing outcomes: winning groups and all trials. Note: The bars represent the number of CPEs coded as “present” across comparisons. N = Total number of effect estimates for each comparison group. * p < .05 ** p < .01 *** p < .001
maladaptive responses to emotional distress, mindfulness, cognitive skills training, and lifestyle. For internalizing outcomes, large differences between winning and non-winning interventions were found for the elements self-exploration or monitoring of thoughts and feelings, training in behavior regulation, mindfulness, awareness, and relaxation, organizational skills, and lifestyle. A moderate-to-large difference was found for emotion recognition and differentiation.

Main Effects: Pooled Effect Sizes

A random-effects three-level model resulted in positive and small-to-moderate effects for externalizing outcomes ($d = 0.313$, $m = 24$, $k = 110$, 95% CI [0.161, 0.465], $p < .001$). This suggests that the interventions improved externalizing outcomes compared to control conditions. For internalizing outcomes, a random-effects three-level model also resulted in a positive and small-to-moderate intervention effect ($d = 0.275$, $m = 13$, $k = 38$, 95% CI [0.073, 0.477], $p = .009$).

There was significant heterogeneity in ES across studies of externalizing outcomes ($Q = 246.94$, $p < .001$). Within-study sampling variance accounted for 43%, 3% attributable to within-study variance, and 54% to between-studies variance. There was also significant heterogeneity in ES for internalizing outcomes ($Q = 52.80$, $p = .044$). Of the total variance, 50% was due to within-study sampling variance, 0% to within-study variance, and 50% to between-studies variance. The modified Egger’s Test (Egger et al., 1997; Marengo & Montag, 2020) was consistent with symmetry in the funnel plot for internalizing outcomes ($B = -1.14$ [-3.81 1.51], $SE = 1.31$, $t(36) = -.87$, $p = .388$). However, for externalizing outcomes, the modified Egger’s Test pointed to a significant positive relation of the ES with the size of the studies ($B = 1.33$ [.04 2.61], $SE = .65$, $t(108) = 2.05$, $p = .043$), suggesting a potential publication bias or small study effect (Borenstein, 2019).

Moderator Analyses

None of the preregistered moderators pertaining to demographic makeup of the studies (i.e., gender composition, type of informant, type of control group) were significantly associated with variation in the effect of treatment programs for internalizing or externalizing outcomes. We performed one additional moderator analysis, for publication year, due to the large variation on this parameter. Publication year was not significantly associated with overall effects.

Table 2 shows the results of the moderator analyses, in which each CPE was applied as a moderator for externalizing symptoms. Two of the CPEs were significantly associated with the post-intervention ES, and two approached significance. Significantly higher intervention effects occurred for the interventions that included the following CPEs: emotional recognition and differentiation
(difference = 0.325, 95% CI [0.047, 0.602], \( p = .022 \)) and psychoeducation (difference = 0.314, 95% CI [0.035, 0.594], \( p = .028 \)). The presence of the following CPEs significantly predicted marginally larger intervention effects: preventing maladaptive responses to emotional distress (difference = 0.323, 95% CI [−0.059, 0.706], \( p = .097 \)) and organizational skills (difference = 0.271, 95% CI [−0.025, 0.567], \( p = .073 \)). None of the CPEs emerged as significant moderators of effects for internalizing outcomes (see, Table 3).

Common Practice Elements Significantly Associated with Effect

Emotional recognition and differentiation was coded as present in nine studies (Bunford, 2017; Coleman et al., 1992; Ellmann, 2002; Goldstein et al., 2018; Hawkins et al., 1991; Jackson, 1991; Jarden, 1995; Ladley, 1997; Schuurmans et al., 2018), 21 ESs, and comprised the following practices: a) Teaching how to recognize triggers for alarm reactions/negative affect; b) Enhancing emotion recognition and/or insight into emotions; c) Discussion of challenging emotional situations; d) Awareness of emotions at physiological level; and e) Recognizing others’ emotions. In five of the eight studies in which emotion recognition and differentiation was present, the targeted outcome was aggression. The three remaining trials targeted interpersonal skills (Bunford, 2017), problem-solving skills (Hawkins et al., 1991), and anxiety in addition to externalizing problems (Schuurmans et al., 2018).

Psychoeducation was coded as present in eight studies (Bunford, 2017; Cianciolo, 1989; Ellmann, 2002; Goldstein et al., 2018; Jackson, 1991; Jewell & Elliff, 2013; Ladley, 1997; Schlichter & Horan, 1981), yielding 18 ESs. The element comprised psychoeducative practices within the following domains: a) Functional and adaptive emotions; b) Problem management; c) Self-esteem and self-worth; d) Stress; e) Anger and aggression; and f) Unspecified psychoeducation (i.e., practices described as psychoeducation about the treatment under scrutiny in the article or manual).

<table>
<thead>
<tr>
<th>Practice Element</th>
<th>Practice Element Present (k)</th>
<th>Practice Element Absent (k)</th>
<th>Mean Difference</th>
<th>ES with Practice Element</th>
<th>ES without Practice Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Emotion recognition and differentiation</td>
<td>64</td>
<td>46</td>
<td>.325*</td>
<td>.464</td>
<td>.140</td>
</tr>
<tr>
<td>2 Preventing maladaptive behavioral responses to emotional distress</td>
<td>97</td>
<td>13</td>
<td>.323†</td>
<td>.371</td>
<td>.048</td>
</tr>
<tr>
<td>3 Self-exploration/monitoring of thoughts and feelings</td>
<td>60</td>
<td>50</td>
<td>.075</td>
<td>.357</td>
<td>.282</td>
</tr>
<tr>
<td>4 Training in behavior regulation</td>
<td>29</td>
<td>81</td>
<td>.269</td>
<td>.522</td>
<td>.253</td>
</tr>
<tr>
<td>5 Mindfulness</td>
<td>35</td>
<td>75</td>
<td>−.005</td>
<td>.309</td>
<td>.314</td>
</tr>
<tr>
<td>6 Training in cognitive skills</td>
<td>79</td>
<td>31</td>
<td>.044</td>
<td>.326</td>
<td>.282</td>
</tr>
<tr>
<td>7 Psychoeducation</td>
<td>71</td>
<td>39</td>
<td>.314*</td>
<td>.469</td>
<td>.155</td>
</tr>
<tr>
<td>8 Problem solving skills</td>
<td>84</td>
<td>26</td>
<td>−.086</td>
<td>.291</td>
<td>.377</td>
</tr>
<tr>
<td>9 Social skills training</td>
<td>39</td>
<td>71</td>
<td>−.143</td>
<td>.226</td>
<td>.368</td>
</tr>
<tr>
<td>10 Organizational skills</td>
<td>55</td>
<td>55</td>
<td>.271†</td>
<td>.460</td>
<td>.189</td>
</tr>
<tr>
<td>11 Lifestyle</td>
<td>19</td>
<td>91</td>
<td>−.008</td>
<td>.246</td>
<td>.334</td>
</tr>
</tbody>
</table>

Note. Differences in effect sizes between programs with or without the practice element were significant at the level of *\( p < .05 \); †\( p < .10 \). \( k \) = number of comparisons. ES = effect size (Cohen’s \( d \).
Table 3. Practice elements associated with stronger or weaker effects on internalizing symptoms.

<table>
<thead>
<tr>
<th>Practice Element</th>
<th>Practice Element Present ($k$)</th>
<th>Practice Element Absent ($k$)</th>
<th>Mean Difference</th>
<th>ES with Practice Element</th>
<th>ES without Practice Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Emotion recognition and differentiation</td>
<td>11</td>
<td>27</td>
<td>−.133</td>
<td>.182</td>
<td>.315</td>
</tr>
<tr>
<td>2 Preventing maladaptive responses to emotional distress</td>
<td>14</td>
<td>24</td>
<td>.052</td>
<td>.310</td>
<td>.259</td>
</tr>
<tr>
<td>3 Self-exploration/monitoring of thoughts and feelings</td>
<td>9</td>
<td>29</td>
<td>.157</td>
<td>.402</td>
<td>.245</td>
</tr>
<tr>
<td>4 Training in behavior regulation</td>
<td>8</td>
<td>30</td>
<td>.197</td>
<td>.438</td>
<td>.241</td>
</tr>
<tr>
<td>5 Mindfulness</td>
<td>12</td>
<td>26</td>
<td>.068</td>
<td>.326</td>
<td>.258</td>
</tr>
<tr>
<td>6 Cognitive skills training</td>
<td>23</td>
<td>15</td>
<td>−.034</td>
<td>.263</td>
<td>.297</td>
</tr>
<tr>
<td>8 Problem solving skills</td>
<td>17</td>
<td>21</td>
<td>−.086</td>
<td>.227</td>
<td>.312</td>
</tr>
<tr>
<td>9 Social skills training</td>
<td>16</td>
<td>22</td>
<td>−.202</td>
<td>.150</td>
<td>.351</td>
</tr>
<tr>
<td>10 Organizational skills</td>
<td>10</td>
<td>28</td>
<td>.130</td>
<td>.377</td>
<td>.247</td>
</tr>
<tr>
<td>11 Lifestyle</td>
<td>9</td>
<td>29</td>
<td>−.255</td>
<td>.083</td>
<td>.338</td>
</tr>
</tbody>
</table>

Note. Differences in effect sizes between programs with or without the practice element were significant at the level of *$p < .05$; † $p < .10$. $k$ = number of comparisons. ES = effect size (Cohen’s $d$). Element 7 did not fulfil the criteria to be defined as a common element for internalizing outcomes.

Discussion

In this study, we have identified practice elements that are common for treatment programs targeting externalizing and/or internalizing symptoms in adolescents in residential settings. We employed two approaches to investigate CPEs across studies and outcomes across residential settings providing treatment programs. By defining winning groups, we enabled identification of elements that are common for effective interventions. Meta-analyses of all ESs within all primary studies yielded estimates of the association of each element with the overall treatment effects across trials. There were considerable discrepancies between which CPEs were associated with beneficial effects across all studies and those identified across winning groups only. The results of the meta-analysis revealed that the effect of treatment programs targeting externalizing symptoms was significantly higher when either of the two CPEs emotion recognition and differentiation, or psychoeducation were present. No CPEs were significantly associated with variance in effects of programs targeting internalizing outcomes. The results of the moderator analyses showed no significant associations between ES and demographic makeup or type of informant. This is in line with the results of a recent study of 742 reports on internalizing and externalizing problems in adolescents in a residential youth care setting (Gevers et al., 2021)

Winning Groups and Common Elements Associated with Effect

The three-level meta-analysis incorporated heterogeneity of effects, the subsequent results reflecting the true effects and not only that of the winning groups. If a winning groups-approach had been conducted alone, the results would have indicated that the two elements training in behavior regulation and mindfulness, awareness, and relaxation were promising elements for both
outcomes. The results of the moderator analyses, however, do not indicate that these two elements contribute significantly to beneficial treatment effects. The combined results illustrate that identifying CPEs in winning groups only, may provide a skewed portrayal of which practices are most promising for building effective interventions. Moreover, the element organizational skills would not have been identified if we had included winning groups only.

While our analyses did not reveal any iatrogenic effects, previous meta-analyses of CPEs in psychotherapy have. For instance, rule setting and parental problem-solving skills have been associated with reduced effects of parenting programs (Leijten et al., 2019).

Identification of CPEs that contribute negatively to the effect of treatment is an important attribute of the meta-analytic approach. The results of our two approaches illustrate the difference between defining CPEs by investigating the most successful programs and investigating all available studies. Opening the ‘black box’ of effective branded programs and revealing CPEs allows for comparison of a broad range of treatments. Reviewing the content and outcomes of therapeutic practices that differ on the surface (i.e., brand names, or therapeutic traditions) enhances the potential for generalizability of the results. In turn, this approach may identify elements that are suitable for a range of populations, settings, and/or disorders (Boustani et al., 2015; Chorpita et al., 2005). Drawing conclusions about CPEs based on both approaches might enhance their theoretical and statistical salience. For instance, the element preventing maladaptive responses to emotional distress had near-significant associations with treatment effect on externalizing outcomes. The element was present in 97% of winning groups for externalizing outcomes. This may be considered an additional indication that this could be a promising element for inclusion in future trials targeting externalizing outcomes. However, our findings suggest that identifying CPEs across effective treatment groups should be used to supplement analyses that take all relevant studies into account.

**Two Promising Common Practice Elements**

The two CPEs that were associated with reduction in externalizing behavior were Emotion Recognition and Differentiation, and Psychoeducation. Emotion differentiation has been defined as the ability to describe and label one’s own emotions in a differentiated and specific manner (Erbas et al., 2019). Emotion recognition and emotion differentiation are related concepts in that the ability to identify and express one’s own emotions is associated with recognizing it in others (Erbas et al., 2016). There is some evidence that this association may predict empathic responses toward others (Israelashvili et al., 2019). Enhanced emotion recognition and emotion differentiation is associated with increased wellbeing, and reduction in stress and emotional pathology (Erbas et al., 2014;
Israelashvili et al., 2019). The two related concepts of emotion recognition and differentiation dovetails onto key elements of Emotion Focused Therapy (EFT) in the emphasis on recognizing one’s own and others’ emotional expressions, and differentiating between primary and secondary emotions (Greenberg, 2004). Emotion recognition and differentiation comprised highly experiential practices (as described in the Results section and Supplementary Table 1) that were initiated through role-play and/or skills training. The element therefore also appears to be closely related to emotion regulation skills in Dialectical Behavior Therapy (DBT). For instance, the DBT skill opposite action is described as: “determining that an emotion either is not warranted by the situation (i.e., “unjustified”) or interferes with effective behavior, while being exposed to emotionally evocative cue or stimuli . . .” (Lynch et al., 2006, p. 471). Our findings therefore suggest that this element, delivered as role-play and/or skills training, could be a promising component to include in future interventions.

Psychoeducation has been defined as “a professionally delivered treatment modality that integrates and synergizes psychotherapeutic and educational interventions” (Lukens & McFarlane, 2004). A recent systematic review and meta-analysis of 276 mental health promotion studies presented psychoeducative interventions as the most effective for improving cognitive skills and mental health literacy among young people (Salazar de Pablo et al., 2020). Psychoeducation has previously been identified as the most common practice element across universal mental health programs for children (Boustani et al., 2020), and prevention programs for university students (Rith-Najarian et al., 2019). The CPE psychoeducation about services has been identified as a primary candidate in promoting engagement in mental health treatment for youth and families (Becker et al., 2015). Psychoeducation has been recommended for use in group therapy settings (Baourda et al., 2022). Psychoeducative interventions have shown robust beneficial effects on mental health problems in adolescents, such as depression (Bevan Jones et al., 2018) and stress and anxiety (Dolan et al., 2021). Psychoeducation is suitable for brief psychotherapies, and for groups, in a variety of settings. Providing it requires modest amounts of formalized training and is therefore an affordable and feasible component for many therapeutic practices. The robust evidence on the efficacy and general utility of psychoeducation, in combination with our findings, suggest that this element should be subject for inclusion and further investigation in future interventions for service-involved youth.

**Internalizing and Externalizing Outcomes**

None of the identified CPEs were associated with significant differences in effect for both internalizing and externalizing outcomes. Effect estimates from the primary studies were skewed toward externalizing outcomes (110 ESs for
externalizing, 39 ESs for internalizing). It is problematic that treatments of externalizing problems in adolescents do not sufficiently cater to the high rates of co-occurring internalizing problems (Granic, 2014). Particularly salient associations have been reported between anxiety and relational and reactive aggression (Chung et al., 2019; Marsee, et al., 2007), indicating several potential causal relationships between the two (Granic, 2014). Our sample may reflect a lack of transdiagnostic treatments in that only eight studies reported on both outcomes. While 13 studies were conducted in correctional facilities, which are more likely to focus on externalizing behaviors, two of them (Macmahon & Gross, 1988; Ahrens & Rexford, 2002) reported on internalizing outcomes only. Conversely, four studies conducted in child welfare or mental health facilities (Coleman et al., 1992; Ellmann, 2003; Jackson, 1991; Ladley, 1997) reported on externalizing outcomes only. In our sample, some adolescents were placed in forensic facilities for truancy (e.g., Stein et al., 2011), theft, or drug charges (e.g., Ahrens & Rexford, 2002). While others, presenting with severe behavioral disorders were placed in “secure” mental health facilities (Schuurmans et al.). Six studies conducted in forensic settings reported high rates of severe mental disorders (Bunford, 2017; Jewell & Elliff, 2013; Macmahon & Gross, 1988; Mitchell et al., 2011; Stein et al., 2011; Ahrens & Rexford, 2002). In one study, juvenile delinquents reported high rates of suicidal ideation, while diagnoses were not reported (Rohde et al., 2004). Our sample shows that delineations of residential settings might not reflect the most important characteristics of service-involved youth with internalizing and externalizing symptoms. For reviewing CPEs, comparing programs targeting adolescent’s disorders across residential settings could be meaningful, while we would advise against such comparisons in general. However, the small and skewed sample may have contributed to the lack of significant moderator effects on internalizing outcomes. It may also have hindered identification of transdiagnostic elements (i.e., elements that predict changes in effects for both outcome categories). Elements with transdiagnostic effects have the highest potential to counteract diverse and co-occurring mental health problems in adolescents (Marchette & Weisz, 2017), and perhaps particularly promising for those in residential settings (Lee & McMillen, 2017).

**Beyond the Multilevel Meta-Analysis**

Several restrictions were applied to the search strategy for this study to accommodate the requirements for inclusion in a multilevel meta-analysis (i.e., reported data on randomized and control groups, ESs, and estimates of variance). When freely reviewing the current literature, a broader picture of therapeutic interventions for service-involved youth in residential settings emerges. Several promising intervention programs exist, many of which are employed in the nexus of the three service domains (i.e., child welfare,
juvenile justice, and mental health care). Notably, Multisystemic Therapy (MST) programs (e.g., Rovers et al., 2019; Rowland et al., 2005), Dialectical Behavior Therapy (DBT) programs (e.g., Little et al., 2010; McCredie et al., 2017), the Children and Residential Experiences (CARE; e.g., Holden et al., 2010), and the Teaching Family Model (TFM; e.g., Fixsen & Blase, 2018; Masuda et al., 2017). Blankenstein et al. (2022) have made an important contribution by exploring effects of the combinations of systemic interventions and secure residential youth care, noting the potential gains of identifying associations between program components and outcomes. James (2011) reviewed the evidence for five group care models, concluding that there is a need for rigorous novel studies that provide detailed reports on treatment as usual, treatment components, and outcomes of relevance for evaluating the efficacy of programs. We could not identify any published studies of CARE or TFM that fit our inclusion criteria of residential care programs for adolescents targeting internalizing or externalizing outcomes, with an RCT design (despite the impressive history of systematic replication and implementation of TFM). For instance, Izzo et al. (2006) reported externalizing outcomes and usage of the CARE program but with a multiple baseline-design. The same goes for DBT and Trauma-Focused DBT, which have been recommended for experimental investigations in residential settings (James et al., 2015). For instance, Sunseri (2004), describes a pre/post-design of a DBT intervention. While the study population, setting, and outcomes matched our criteria, the study did not meet the requirements for the study design. If future efforts of identifying CPEs and their associations with outcomes are to succeed, trials with a randomized design that adhere to reporting standards such as those proposed by Lee and Barth (2011), must be conducted.

Strengths and Limitations

A main strength of this study is that CPEs were analyzed in a three-level model. The method corrects for data dependency (nesting) in all levels of analysis. It allows for inclusion of all outcome dependency measures, and for differential estimates of ESs for when an element is present versus absent. With the methods currently available, these estimates are the closest we can get to unraveling associations between practices and treatment effects. The multilevel meta-analytic approach is therefore superior to that of significance tests of pooled effect sizes, or that of identifying CPEs of effective programs. We offer results that are applicable for determining which practices to include in interventions for adolescents, drawing on evidence from a variety of residential settings. The results may also be used for identifying which practices are commonly used but not associated with beneficial effects. Reporting which elements were commonly found in effective programs, sheds additional light on which elements are most promising, and
highlights the discrepancies between the two approaches. Importantly, the moderator analyses do not allow for causal claims about effects of CPEs. The results are only appropriately used in generating hypotheses about practices to include in future trials investigating treatment effects.

The current study has several limitations. Our sample included treatment settings that differ substantially with regard to treatment providers, intensity, and duration. Only nine primary studies reported follow-up measurements (ranging between two weeks and eleven months), which makes it impossible to make assumptions about the duration of treatment effects. Diverse measurements were used for both internalizing and externalizing outcomes, reducing the reliability of comparisons. This is a common problem when conducting meta-analyses of treatment effects. While cohort studies, qualitative, naturalistic, and observational studies reporting on residential programs for youth exist, only RCTs were included for comparison in the present meta-analysis. As a result, the number of included primary studies was small. Ten primary studies were published over 20 years ago. This inevitably reduces the generalizability and relevance of our findings and illustrates the tradeoffs inherent in a multilevel model approach. For the multilevel analysis to establish associations between CPEs and ESs, primary studies must meet stringent requirements for how and which data are reported. As noted by Lee et al. (2014), non-randomized studies may be eligible for inclusion when identifying CPEs, as the RCT literature is often limited (especially for behavioral interventions). The multilevel meta-analysis takes us a step further in establishing associations between elements and effects. Unfortunately, the results only reflect the fraction of studies that employ a RCT design. A recent meta-analytic review of outcomes of interventions targeting internalizing problems in juvenile justice facilities identified only eleven studies, six of which were non-randomized (Kumm et al., 2019). The small sample and low methodological quality of primary studies rendered the results ambiguous. This reinforces our claim that novel high-quality RCT studies are needed. While the multilevel meta-analytic approach has its advantages, we welcome future efforts employing alternate strategies to investigate common elements, including those pertaining to delivery processes, implementation, and fidelity.

There was significant heterogeneity of effects for both outcomes, largely attributable to within-study sampling variance and between-study variance. The between-study variance was likely associated with the variability of outcome measurements. For externalizing outcomes, the results of the modified Egger’s Test combined with the large proportion of small-scale RCTs (mean randomized sample size N = 77) indicate a possible small-study publication bias. This may reflect the considerable proportion of studies published as part of doctoral dissertations (N = 8). The randomized sample size was used for one small study (N = 27) because post-sample sizes were not reported. Finally, meta-
analyses of associations between program elements and program effects cannot rule out the possibility of confounders driving the effect of the element (e.g., more rigorous therapist training, higher program fidelity; Leijten et al., 2021).

**Protocol Deviations**

This study deviated from the preregistered study protocol on two accounts. We did not conduct exploratory analyses for discrete practices to avoid the large number of comparisons needed to conduct analyses for an additional level of 58 potential moderators (practices). We did not investigate secondary outcomes pertaining to quality of life or social functioning, as few and disparate measurements were reported for these outcomes.

**Conclusion**

We have identified two CPEs that are associated with beneficial effects of therapeutic interventions for adolescents with externalizing problems across a wide variety of residential settings: *emotion recognition and differentiation*, and *psychoeducation*. The utility of our findings are as follows: The two elements should be subject to further investigation in trials with factorial or stepwise design, micro- or dismantling trials. By contrasting two approaches to identifying CPEs, our analyses highlight the importance of including all relevant studies and analyzing whether CPEs are associated with treatment effects.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

This work was supported by the Norges Forskningsråd [262789,288083].

**ORCID**

Line Solheim Kvamme [http://orcid.org/0000-0002-2946-1148](http://orcid.org/0000-0002-2946-1148)
Serap Keles [http://orcid.org/0000-0002-1174-0667](http://orcid.org/0000-0002-1174-0667)
Ragnhild Bang Nes [http://orcid.org/0000-0001-7828-7863](http://orcid.org/0000-0001-7828-7863)
Linda Vaskinn [http://orcid.org/0000-0002-6505-9519](http://orcid.org/0000-0002-6505-9519)
Pamela M. Waaler [http://orcid.org/0000-0002-0480-1302](http://orcid.org/0000-0002-0480-1302)
Tore Wentzel-Larsen [http://orcid.org/0000-0002-0318-4162](http://orcid.org/0000-0002-0318-4162)
John Kjøbli [http://orcid.org/0000-0002-5639-1691](http://orcid.org/0000-0002-5639-1691)
References


Ahrens, J., & Rexford, L. (2002). Cognitive processing therapy for incarcerated adolescents with PTSD. *Journal of Aggression, Maltreatment & Trauma, 6*(1), 201–216. https://doi.org/10.1300/J146v06n01_10


