The impact of enhancing students’ social and emotional learning: A meta-analysis of school-based universal interventions

Joseph Durlak, Roger P. Weissberg, Allison Dymnicki, Rebecca Taylor, & Kriston Schellinger


For more information, please contact Dr. Roger P. Weissberg at the address and telephone number below.
This article presents findings from a meta-analysis of 213 school-based, universal social and emotional learning (SEL) programs involving 270,034 kindergarten through high school students. Compared to controls, SEL participants demonstrated significantly improved social and emotional skills, attitudes, behavior, and academic performance that reflected an 11-percentile-point gain in achievement. School teaching staff successfully conducted SEL programs. The use of 4 recommended practices for developing skills and the presence of implementation problems moderated program outcomes. The findings add to the growing empirical evidence regarding the positive impact of SEL programs. Policy makers, educators, and the public can contribute to healthy development of children by supporting the incorporation of evidence-based SEL programming into standard educational practice.
There is broad agreement among educators, policy makers, and the public that educational systems should graduate students who are proficient in core academic subjects, able to work well with others from diverse backgrounds in socially and emotionally skilled ways, practice healthy behaviors, and behave responsibly and respectfully (Association for Supervision and Curriculum Development, 2007; Greenberg et al., 2003). In other words, schools have an important role to play in raising healthy children by fostering not only their cognitive development but also their social and emotional development. Yet schools have limited resources to address all of these areas and are experiencing intense pressures to enhance academic performance. Given time constraints and competing demands, educators must prioritize and effectively implement evidence-based approaches that produce multiple benefits.

It has been posited that universal school-based efforts to promote students’ social and emotional learning (SEL) represent a promising approach to enhance children’s success in school and life (Elias et al., 1997; Zins & Elias, 2006). Extensive developmental research indicates that effective mastery of social-emotional competencies is associated with greater well-being and better school performance whereas the failure to achieve competence in these areas can lead to a variety of personal, social, and academic difficulties (Eisenberg, 2006; Guerra & Bradshaw, 2008; Masten & Coatsworth, 1998; Weissberg & Greenberg, 1998). The findings from various clinical, prevention, and youth development studies have stimulated the creation of many school-based interventions specifically designed to promote young people’s SEL (Greenberg et al., 2003). On the other hand, several researchers have questioned the extent to which promoting children’s social and emotional skills will actually improve their behavioral and academic outcomes (Duncan et al., 2007; Zeidner, Roberts, & Matthews, 2002). This meta-analysis examines the effects of school-based SEL programming on children’s behaviors and academic performance, and discusses the implications of these findings for educational policies and practice.

What Is Social and Emotional Learning?

The SEL approach integrates competence promotion and youth development frameworks for reducing risk factors and fostering protective mechanisms for positive adjustment (Benson, 2006; Catalano, Berghlund, Ryan, Lonczak, & Hawkins, 2002; Guerra & Bradshaw, 2008; Weissberg, Kumpfer, & Seligman, 2003). SEL researchers and program designers build from Waters and Sroufe’s (1983) description of competent people as those who have the abilities “to generate and coordinate flexible, adaptive responses to demands and to generate and capitalize on opportunities in the environment” (p. 80). Elias et al. (1997) defined SEL as the process of acquiring core competencies to recognize and manage emotions, set and achieve positive goals, appreciate the perspectives of others, establish and maintain positive relationships, make responsible decisions, and handle interpersonal situations constructively. The proximal goals of SEL programs are to foster the development of five interrelated sets of cognitive, affective, and behavioral competencies: self-awareness, self-management, social awareness, relationship skills, and responsible decision making (Collaborative for Academic, Social, and Emotional Learning, 2005). These competencies, in turn, should provide a foundation for better adjustment and academic performance as reflected in more positive social behaviors, fewer conduct problems, less emotional distress, and improved test scores and grades (Greenberg et al., 2003). Over time, mastering SEL competencies results in a developmental progression that leads to a shift from being predominantly controlled by external factors to acting increasingly in accord with internalized beliefs and values, caring and concern for others, making good decisions, and taking responsibility for one’s choices and behaviors (Bear & Watkins, 2006).

Within school contexts, SEL programming incorporates two coordinated sets of educational strategies to enhance school performance and youth development (Collaborative for Academic, Social, and Emotional Learning, 2005). The first involves instruction in processing, integrating, and selectively applying social and emotional skills in developmentally, contextually, and culturally appropriate ways (Crick & Dodge, 1994; Izard, 2002; Lemerise & Arsenio, 2000). Through systematic instruction, SEL skills may be taught, modeled, practiced, and applied to diverse situations so that students use them as part of their daily repertoire of behaviors (Ladd & Mize, 1983; Weissberg, Caplan, & Sivo, 1989). In addition, many programs help students apply SEL skills in preventing specific problem behaviors such as substance use, interpersonal violence, bullying, and school failure (Zins & Elias, 2006). Quality SEL instruction also provides students with opportunities to contribute to their class, school, and community and experience the satisfaction, sense of belonging, and enhanced motivation.
that comes from such involvement (Hawkins, Smith, & Catalano, 2004). Second, SEL programming fosters students’ social-emotional development through establishing safe, caring learning environments involving peer and family initiatives, improved classroom management and teaching practices, and whole-school community-building activities (Cook et al., 1999; Hawkins et al., 2004; Schaps, Battistich, & Solomon, 2004). Together these components promote personal and environmental resources so that students feel valued, experience greater intrinsic motivation to achieve, and develop a broadly applicable set of social-emotional competencies that mediate better academic performance, health-promoting behavior, and citizenship (Greenberg et al., 2003).

Recent Relevant Research Reviews

During the past dozen years there have been many informative research syntheses of school-based prevention and promotion programming. These reviews typically include some school-based, universal SEL program evaluations along with an array of other interventions that target the following outcomes: academic performance (Wang, Haertel, & Walberg, 1997; Zins et al., 2004), antisocial and aggressive behavior (Lösel & Beelman, 2003; Wilson & Lipsey, 2007), depressive symptoms (Horowitz & Garber, 2006), drug use (Tobler et al., 2000), mental health (Durlak & Wells, 1997; Greenberg, Domitrovich, & Bumbarger, 2001), problem behaviors (Wilson, Gottfredson, & Najaka, 2001), or positive youth development (Catalano et al., 2002). Although these reports differ substantially in terms of which intervention strategies, student populations, and behavioral outcomes are examined, they have reached a similar conclusion that universal school-based interventions are generally effective. However, no review to date has focused exclusively on SEL programs to examine their impact across diverse student outcomes.

The Current Meta-Analysis: Research Questions and Hypotheses

This paper reports on the first large-scale meta-analysis of school-based programs to promote students’ social and emotional development. In contrast to most previous reviews that focus on one major outcome (e.g., substance abuse, aggression, academic performance), we explored the effects of SEL programming across multiple outcomes: social and emotional skills, attitudes toward self and others, positive social behavior, conduct problems, emotional distress, and academic performance. Moreover, we were interested in interventions for the entire student body (universal interventions) and thus did not examine programs for indicated populations, that is, for students already demonstrating adjustment problems. These latter programs have been evaluated in a separate report (Payton et al., 2008).

The proliferation of new competence-promotion approaches led to several important research questions about school-based interventions to foster students’ social and emotional development. For example, what outcomes are achieved by interventions that attempt to enhance children’s emotional and social skills? Can SEL interventions promote positive outcomes and prevent future problems? Can programs be successfully conducted in the school setting by existing school personnel? What variables moderate the impact of school-based SEL programs? Next, we address these questions and offer hypotheses about expected findings.

The findings from several individual studies and narrative reviews indicate that SEL programs are associated with positive results such as improved attitudes about the self and others, increased prosocial behavior, lower levels of problem behaviors and emotional distress, and improved academic performance (Catalano et al., 2002; Greenberg et al., 2003; Zins et al., 2004). Thus, our first hypothesis was that our meta-analysis of school-based SEL programs would yield significant positive mean effects across a variety of skill, attitudinal, behavioral, and academic outcomes (Hypothesis 1).

Ultimately, interventions are unlikely to have much practical utility or gain widespread acceptance unless they are effective under real-world conditions. Thus, we investigated whether SEL programs can be incorporated into routine educational practice; that is, can they be successfully delivered by existing school staff during the regular school day? In our analyses, we separated interventions conducted by regular school staff and those administered by nonschool personnel (e.g., university researchers, outside consultants). We predicted that programs conducted by classroom teachers and other school staff would produce significant outcomes (Hypothesis 2).

Many school-based SEL programs involve the delivery of classroom curricula designed to promote social-emotional competencies in developmentally and culturally appropriate ways (Collaborative for Academic, Social, and Emotional Learning, 2005). There are also multicomponent programs that
supplement classroom programming with school-wide components (Greenberg et al., 2003). We expected that interventions that combined components within and outside of the daily classroom routine would yield stronger effects than those that were only classroom based (Hypothesis 3). This expectation is grounded in the premise that the broader ecological focus of multicomponent programs that extend beyond the classroom should better support and sustain new skill development (Tolan, Guerra, & Kendall, 1995).

We also predicted that two key variables would moderate student outcomes: the use of recommended practices for developing skills and adequate program implementation. Extensive research in school, community, and clinical settings has led several authors to offer recommendations on what procedures should be followed for effective skill training. For example, there is broad agreement that programs are likely to be effective if they use a sequenced step-by-step training approach, use active forms of learning, focus sufficient time on skill development, and have explicit learning goals (Bond & Hauf, 2004; Durlak, 1997; Dusenbury & Falco, 1995; Gresham, 1995). These four recommended practices form the acronym SAFE (for sequenced, active, focused, and explicit; see the Method section). A meta-analysis of after-school programs that sought to develop personal and social skills found that program staff who followed these four recommended practices were more effective than those who did not follow these procedures (Durlak, Weissberg, & Pachan, 2010). Moreover, the literature suggests that these recommended practices are important in combination with one another rather than as independent factors. In other words, sequenced training will not be as effective unless active forms of learning are used and sufficient time is focused on reaching explicit learning goals. Therefore, we coded how many of the four practices were used in SEL interventions and expected to replicate the previous finding that staff using all four practices would be more successful than those who did not (Hypothesis 4).

For example, new behaviors and more complicated skills usually need to be broken down into smaller steps and sequentially mastered, suggesting the benefit of a coordinated sequence of activities that links the learning steps and provides youth with opportunities to connect these steps (Sequenced). Gresham (1995) has noted that it is “important to help children learn how to combine, chain and sequence behaviors that make up various social skills’” (p. 1023). Lesson plans and program manuals are often used for this purpose.

An effective teaching strategy for many youth emphasizes the importance of active forms of learning that require youth to act on the material (Active). “It is well documented that practice is a necessary condition for skill acquisition” (Salas & Cannon-Bowers, 2001, p. 480). Sufficient time and attention must also be devoted to any task for learning to occur (Focus). Therefore, some time should be set aside primarily for skill development. Finally, clear and specific learning objectives over general ones are preferred because it is important that youth know what they are expected to learn (Explicit).

Finally, there is increasing recognition that effective implementation influences program outcomes (Durlak & Dupre, 2008) and that problems encountered during program implementation can limit the benefits that participants might derive from intervention. Therefore, we hypothesized that SEL programs that encountered problems during program implementation would be less successful than those that did not report such problems (Hypothesis 5).

In sum, this article describes the results of a meta-analysis of school-based universal SEL programs for school children. We hypothesized that (a) SEL programs would yield significant mean effects across skill, attitudinal, behavioral, and academic domains; (b) teachers would be effective in administering these programs; and (c) multicomponent programs would be more effective than single-component programs. We also expected that program outcomes would be moderated by (d) the use of recommended training practices (SAFE practices) and (e) reported implementation problems.

Method

Literature Search

Four search strategies were used in an attempt to secure a systematic, nonbiased, representative sample of published and unpublished studies. First, relevant studies were identified through computer searches of PsycInfo, Medline, and Dissertation Abstracts using the following search terms and their variants: social and emotional learning, competence, assets, health promotion, prevention, positive youth development, social skills, self-esteem, empathy, emotional intelligence, problem solving, conflict resolution, coping, stress reduction, children, adolescents, intervention, students, and schools. Second, the reference lists of each identified study and of reviews of psychosocial interventions for youth were examined.
Third, manual searches were conducted in 11 journals producing relevant studies from January 1, 1970 through December 31, 2007. These were the American Educational Research Journal, American Journal of Community Psychology, Child Development, Journal of Research in Adolescence, Journal of Consulting and Clinical Psychology, Journal of Primary Prevention, Journal of School Psychology, Journal of Youth and Adolescence, Prevention Science, Psychology in the Schools, and School Psychology Review. Fourth, searches were made of organization Web sites promoting youth development and social-emotional learning, and researchers who presented relevant work at national prevention and community conferences were contacted for complete reports. The final study sample has little overlap with previous meta-analyses of school-based preventive interventions. No more than 12% of the studies in any of the previous reviews (Durlak & Wells, 1997; Horowitz & Garber, 2007; Lösel & Beelman, 2003; Tobler et al., 2000; Wilson et al., 2001; Wilson & Lipsey, 2007) were part of our study sample, and 63% of the studies we reviewed were not included in any of these previous reviews. This is due to a number of reasons including (a) 36% of studies in the current review were published in the past decade, (b) previous reviews have focused primarily on negative outcomes and not on positive social-emotional skills and attitudes, and (c) other studies have not included such a broad range of age groups (i.e., kindergarten through high school students).

Inclusion Criteria

Studies eligible for review were (a) written in English; (b) appeared in published or unpublished form by December 31, 2007; (c) emphasized the development of one or more SEL skills; (d) targeted students between the ages of 5 and 18 without any identified adjustment or learning problems; (e) included a control group; and (f) reported sufficient information so that effect sizes (ESs) could be calculated at post and, if follow-up data were collected, at least 6 months following the end of intervention.

Exclusion Criteria

We excluded studies targeting students who had preexisting behavioral, emotional, or academic problems. Additionally, we excluded programs whose primary purpose was to promote achievement through various types of educational curricula, instructional strategies, or other forms of academic assistance, as well as interventions that focused solely on outcomes related to students’ physical health and development (e.g., programs to prevent AIDS, pregnancy, or drug use, or those seeking to develop healthy nutrition and exercise patterns). Finally, we excluded small-group out-of-class programs that were offered during study hall, gym class, or in school after the school day ended. Although some of these programs technically qualify as universal interventions, they differed in several respects from the other reviewed interventions. For example, they did not involve entire classes but were limited to those students who volunteered (thus introducing the possibility of self-selection bias) and they usually had much smaller sample sizes and were briefer in duration.

Dealing With Multiple Cohorts or Multiple Publications on the Same Cohort

Multiple interventions from the same report were coded and analyzed separately if the data related to distinct intervention formats (e.g., classroom versus multicomponent) and contained separate cohorts, or if a single report reported the results for an original cohort and a replication sample. Multiple papers evaluating the same intervention but containing different outcome data at post or follow-up for the same cohort were combined into a single study.

Independent Variable: Intervention Formats

The major independent variables were intervention format, the use of four recommended practices related to skill development (SAFE practices), and reported implementation problems. The intervention format used to promote students’ social and emotional development was categorized in the following three mutually exclusive ways based on the primary change agent and whether multi-component strategies were used to influence students.

Class by teacher. The most common strategy (53% of interventions) involved classroom-based interventions administered by regular classroom teachers (Class by Teacher). These usually took the form of a specific curriculum and set of instructional strategies (e.g., behavior rehearsal, cooperative learning) that sought to develop specific social and emotional skills.

Class by nonschool personnel. These interventions were similar to Class by Teacher approaches with the major difference being that nonschool personnel, such as university researchers or outside consultants, administered the intervention.
Multicomponent programs. These approaches typically had two components and often supplemented teacher-administered classroom interventions with a parent component or schoolwide initiatives. In some projects, parents worked with their child to complete skill-related homework assignments or attended parent discussion and training groups (e.g., Kumpfer, Alvarado, Tait, & Turner, 2002). Others involved schoolwide organizational changes. For example, these efforts might begin with the formation of a planning team that develops new policies and procedures to reorganize school structures and then institutes practices to encourage and support students’ social and emotional development (e.g., Cook, Murphy, & Hunt, 2000; Flay, Allred, & Ordway, 2001; Hawkins et al., 2004).

Potential Moderators of Outcome: SAFE and Implementation

SAFE. Interventions were coded dichotomously (yes or no) according to whether or not each of four recommended practices identified by the acronym SAFE was used to develop students’ skills: (a) Does the program use a connected and coordinated set of activities to achieve their objectives relative to skill development? (Sequenced); (b) Does the program use active forms of learning to help youth learn new skills? (Active); (c) Does the program have at least one component devoted to developing personal or social skills? (Focused); and (d) Does the program target specific SEL skills rather than targeting skills or positive development in general terms? (Explicit). Reports rarely contained data on the extent to which each of the above four practices were used (e.g., how often or to what degree active forms of learning were used) and, therefore, dichotomous coding was necessary. For example, any time spent on active learning (e.g., role playing or behavioral rehearsal) was credited as long as it afforded students the opportunity to practice or rehearse SEL skills. Further details on these practices are available in the coding manual and in Durlak et al. (2010). Programs that followed or failed to follow all four practices were called SAFE and Other programs, respectively.

Program implementation. First, we noted whether authors monitored the process of implementation in any way. If the answer was affirmative, we then coded reports (yes or no) for instances of implementation problems (e.g., when staff failed to conduct certain parts of the intervention or unexpected developments altered the execution of the program). Thus, a program was only coded as having no implementation problems if implementation was monitored and authors reported no problems or that the program was delivered as intended.

Methodological Variables

To assess how methodological features might influence outcomes, three variables were coded dichotomously (randomization to conditions, use of a reliable outcome measure, and use of a valid outcome measure; each as yes or no). An outcome measure’s reliability was considered acceptable if kappa or alpha statistics were ≥ .60, reliability calculated by product moment correlations was ≥ .70, and level of percentage agreement by raters was ≥ .80. A measure was considered valid if the authors cited data confirming the measure’s construct, concurrent, or predictive validity. Reliability and validity were coded dichotomously because exact psychometric data were not always available. Additionally, we coded attrition as a continuous variable in two ways: (a) as total attrition from the combined intervention and control group sample from pre to post and (b) as differential attrition, assessed as the percentage of attrition from the control group subtracted from the attrition percentage of the intervention group.

Dependent Variables: Student Outcomes

The dependent variables used in this meta-analysis were six different student outcomes: (a) social and emotional skills, (b) attitudes toward self and others, (c) positive social behaviors, (d) conduct problems, (e) emotional distress, and (f) academic performance.

Social and emotional skills. This category includes evaluations of different types of cognitive, affective, and social skills related to such areas as identifying emotions from social cues, goal setting, perspective taking, interpersonal problem solving, conflict resolution, and decision making. Skill assessments could be based on the reports from the student, a teacher, a parent, or an independent rater. However, all the outcomes in this category reflected skill acquisition or performance assessed in test situations or structured tasks (e.g., interviews, role plays, or questionnaires). In contrast, teacher ratings of students’ behaviors manifested in daily situations (e.g., a student’s ability to control anger or work well with others) were placed in the positive social behavior category below.
Attitudes toward self and others. This category combines positive attitudes about the self, school, and social topics. It included self-perceptions (e.g., self-esteem, self-concept, and self-efficacy), school bonding (e.g., attitudes toward school and teachers), and conventional (i.e., prosocial) beliefs about violence, helping others, social justice, and drug use. All the outcomes in this category were based on student self-reports. We combined these three outcomes to avoid extremely small cell sizes for subsequent analyses.

Positive social behavior. This category included outcomes such as getting along with others derived from the student, teacher, parent, or an independent observer. These outcomes reflect daily behavior rather than performance in hypothetical situations, which was treated as a social and emotional skill outcome. For example, teacher ratings of social skills drawn from Elliott and Gresham’s Social Skills Rating Scale (Elliott, Gresham, Freeman, & McCloskey, 1988) were put into the positive social behavior outcome category.

Conduct problems. This category included measures of different types of behavior problems, such as disruptive class behavior, noncompliance, aggression, bullying, school suspensions, and delinquent acts. These measures, such as the Child Behavior Checklist (Achenbach, 1991), could also come from student self-reports, teacher or parent ratings, or independent observers, or, in the case of school suspensions, only from school records.

Emotional distress. This category consisted of measures of internalized mental health issues. These included reports of depression, anxiety, stress, or social withdrawal, which could be provided by students, teachers, or parents on measures such as the Children’s Manifest Anxiety Scale (Kitano, 1960).

Academic performance. Academic performance included standardized reading or math achievement test scores from such measures as the Stanford Achievement Test or the Iowa Test of Basic Skills, and school grades in the form of students’ overall GPA or their grades in specific subjects (usually reading or math). Only data drawn from school records were included. Teacher-developed tests, teacher ratings of academic competence, and IQ measures such as the Stanford Binet were not included.

Coding Reliability

A coding system available from the first author was developed to record information about each report such as its date of appearance and source, characteristics of the participants, methodological features, program procedures, and measured outcomes. Trained research assistants working in pairs but at different time periods and on different aspects of the total coding system completed the coding. Reliability of coding was estimated by having pairs of students independently code a randomly selected 25% sample of the studies. Kappa coefficients corrected for chance agreement were acceptable across all codes reported in this review (mean kappa was 0.69). Raters’ agreements on continuous variables were all above 0.90. Any disagreements in coding were eventually resolved through discussion.

Calculation of Effects and General Analytic Strategies

Hedge’s g (Hedges & Olkin, 1985) was the index of effect adjusted whenever possible for any preintervention differences between intervention and control groups (e.g., Wilson & Lipsey, 2007; Wilson et al., 2001). All ESs were calculated such that positive values indicated a favorable result for program students over controls. When means and standard deviations were not available, we used estimation procedures recommended by Lipsey and Wilson (2001). If the only information in the report was that the results were nonsignificant and attempts to contact authors did not elicit further information, the ES was conservatively set at zero. There were 45 imputed zeros among the outcomes, and subsequent analyses indicated these zeros were not more likely to be associated with any coded variables.

One ES per study was calculated for each outcome category. In addition, we corrected each ES for small sample bias, weighted ESs by the inverse of their variance prior to any analysis, and calculated 95% confidence intervals around each mean. When testing our hypotheses, a .05 probability level was used to determine statistical significance. A mean ES is significantly different from zero when its 95% confidence intervals do not include zero. The method of examining overlapping confidence intervals (Cumming & Finch, 2005) was used to determine if the mean ESs from different groups of studies differed significantly. Finally, the method used for all analyses was based on a random effects model using maximum likelihood estimation procedure (Lipsey & Wilson, 2001).

The significance of the heterogeneity of a group of ESs was examined through the Q statistic.
A significant $Q$ value suggests studies are not drawn from a common population whereas a nonsignificant value indicates the opposite. In addition, we used the $I^2$ statistic (Higgins, Thompson, Deeks, & Altman, 2003), which reflects the degree (as opposed to the statistical significance) of heterogeneity among a set of studies along a 0%–100% scale.

**Results**

**Descriptive Characteristics of Reviewed Studies**

The sample consisted of 213 studies that involved 270,034 students. Table 1 summarizes some of the features of these investigations. Most papers (75%) were published during the last two decades. Almost half (47%) of the studies employed randomized designs. More than half the programs (56%) were delivered to elementary school students, just under a third (31%) involved middle school students, and the remainder included high school students. Although nearly one third of the reports contained no information on student ethnicity (31%) or socioeconomic status (32%), several interventions occurred in schools serving a mixed student body in terms of ethnicity (35%) or socioeconomic status (25%). Just under half of the studies were conducted in urban schools (47%). The majority of SEL programs were classroom based, either delivered by teachers (53%) or nonschool personnel (21%), and 26% were multicomponent programs. About 77% of the programs lasted for less than a year, 11% lasted 1–2 years, and 12% lasted more than 2 years.

**SEL Programs Significantly Improve Students’ Skills, Attitudes, and Behaviors**

The grand study-level mean for all 213 interventions was 0.30 (CI = 0.26–0.33), which was statistically significant from zero. The $Q$ value of 2,453 was significant ($p \leq .001$) and the $I^2$ was high (91%), indicating substantial heterogeneity among studies and suggesting the existence of one or more variables that might moderate outcomes.

Table 2 presents the mean effects and their 95% confidence intervals obtained at post across all reviewed programs in each outcome category. All six means (range = 0.22 to 0.57) are significantly greater than zero and confirm our first hypothesis. Results (based on 35–112 interventions depending on the outcome category) indicated that, compared to controls, students demonstrated enhanced SEL

<table>
<thead>
<tr>
<th>General publication features</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955–1979</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>1980–1989</td>
<td>35</td>
<td>16</td>
</tr>
<tr>
<td>1990–1999</td>
<td>83</td>
<td>39</td>
</tr>
<tr>
<td>2000–2007</td>
<td>77</td>
<td>36</td>
</tr>
<tr>
<td>Source of report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Published article/books</td>
<td>172</td>
<td>81</td>
</tr>
<tr>
<td>Unpublished reports</td>
<td>41</td>
<td>19</td>
</tr>
<tr>
<td>Methodological features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>99</td>
<td>47</td>
</tr>
<tr>
<td>No</td>
<td>114</td>
<td>53</td>
</tr>
<tr>
<td>Mean percent of attrition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implementation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not reported on</td>
<td>91</td>
<td>43</td>
</tr>
<tr>
<td>No significant problems</td>
<td>74</td>
<td>35</td>
</tr>
<tr>
<td>Significant problems</td>
<td>48</td>
<td>22</td>
</tr>
<tr>
<td>Use of reliable outcome measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>550</td>
<td>76</td>
</tr>
<tr>
<td>No</td>
<td>176</td>
<td>24</td>
</tr>
<tr>
<td>Use of valid outcome measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>369</td>
<td>51</td>
</tr>
<tr>
<td>No</td>
<td>357</td>
<td>49</td>
</tr>
<tr>
<td>Source of outcome data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child</td>
<td>382</td>
<td>53</td>
</tr>
<tr>
<td>Other (parent, teacher, observer, school records)</td>
<td>422</td>
<td>47</td>
</tr>
<tr>
<td>Participant features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational level of participants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary school (Grades K–5)</td>
<td>120</td>
<td>56</td>
</tr>
<tr>
<td>Middle school (Grades 6–8)</td>
<td>66</td>
<td>31</td>
</tr>
<tr>
<td>High school (Grades 9–12)</td>
<td>27</td>
<td>13</td>
</tr>
<tr>
<td>Intervention features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention format</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class by Teacher</td>
<td>114</td>
<td>53</td>
</tr>
<tr>
<td>Class by Nonschool Personnel</td>
<td>44</td>
<td>21</td>
</tr>
<tr>
<td>Multicomponent</td>
<td>55</td>
<td>26</td>
</tr>
<tr>
<td>Use of recommended training procedures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention rated as SAFE</td>
<td>176</td>
<td>83</td>
</tr>
<tr>
<td>Intervention not rated as SAFE</td>
<td>37</td>
<td>17</td>
</tr>
<tr>
<td>Number of sessions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean number of sessions</td>
<td>40.8</td>
<td>87</td>
</tr>
<tr>
<td>Median number of sessions</td>
<td>24</td>
<td>13</td>
</tr>
<tr>
<td>Locale of intervention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>186</td>
<td>87</td>
</tr>
<tr>
<td>Outside the United States</td>
<td>27</td>
<td>13</td>
</tr>
<tr>
<td>General area of school</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>99</td>
<td>47</td>
</tr>
<tr>
<td>Suburban</td>
<td>35</td>
<td>16</td>
</tr>
<tr>
<td>Rural</td>
<td>31</td>
<td>15</td>
</tr>
<tr>
<td>Combination of areas</td>
<td>30</td>
<td>14</td>
</tr>
<tr>
<td>Did not report</td>
<td>18</td>
<td>8</td>
</tr>
</tbody>
</table>
skills, attitudes, and positive social behaviors following intervention, and also demonstrated fewer conduct problems and had lower levels of emotional distress. Especially noteworthy from an educational policy perspective, academic performance was significantly improved. The overall mean effect did not differ significantly for test scores and grades (mean ESs = 0.27 and 0.33, respectively). Although only a subset of studies collected information on academic performance, these investigations contained large sample sizes and involved a total of 135,396 students.

Follow-Up Effects

Thirty-three of the studies (15%) met the criteria of collecting follow-up data at least 6 months after the intervention ended. The average follow-up period across all outcomes for these 33 studies was 92 weeks (median = 52 weeks; means range from 66 weeks for SEL skills to 150 weeks for academic performance). The mean follow-up ESs remained significant for all outcomes in spite of reduced numbers of studies assessing each outcome: SEL skills (ES = 0.26; k = 8), attitudes (ES = 0.11; k = 16), positive social behavior (ES = 0.17; k = 12), conduct problems (ES = 0.14; k = 21), emotional distress (ES = 0.15; k = 11), and academic performance (ES = 0.32; k = 8). Given the limited number of follow-up studies, all subsequent analyses were conducted at post only.

### Table 2

Mean Effects and .05 Confidence Intervals at Post for Total Sample and Each Intervention Format

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>SEL skills</th>
<th>Attitudes</th>
<th>Positive social behavior</th>
<th>Conduct problems</th>
<th>Emotional distress</th>
<th>Academic performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total sample</td>
<td>ES: 0.57*</td>
<td>0.23*</td>
<td>0.24*</td>
<td>0.22*</td>
<td>0.24*</td>
<td>0.27*</td>
</tr>
<tr>
<td></td>
<td>CI: 0.48 to 0.67</td>
<td>0.16 to 0.30</td>
<td>0.16 to 0.32</td>
<td>0.16 to 0.29</td>
<td>0.14 to 0.35</td>
<td>0.15 to 0.39</td>
</tr>
<tr>
<td></td>
<td>N: 68</td>
<td>106</td>
<td>86</td>
<td>112</td>
<td>49</td>
<td>35</td>
</tr>
<tr>
<td>Class by Teacher</td>
<td>ES: 0.62*</td>
<td>0.23*</td>
<td>0.26*</td>
<td>0.20*</td>
<td>0.25*</td>
<td>0.34*</td>
</tr>
<tr>
<td></td>
<td>CI: 0.41 to 0.82</td>
<td>0.17 to 0.29</td>
<td>0.15 to 0.38</td>
<td>0.12 to 0.29</td>
<td>0.08 to 0.43</td>
<td>0.16 to 0.52</td>
</tr>
<tr>
<td></td>
<td>N: 40</td>
<td>59</td>
<td>59</td>
<td>53</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Class by Nonschool Personnel</td>
<td>ES: 0.87*</td>
<td>0.14*</td>
<td>0.23</td>
<td>0.17*</td>
<td>0.21</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>CI: 0.58 to 1.16</td>
<td>0.02 to 0.25</td>
<td>−0.04 to 0.50</td>
<td>0.02 to 0.33</td>
<td>−0.01 to 0.43</td>
<td>−0.19 to 0.43</td>
</tr>
<tr>
<td></td>
<td>N: 21</td>
<td>18</td>
<td>11</td>
<td>16</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>Multicomponent</td>
<td>ES: 0.12</td>
<td>0.23*</td>
<td>0.19</td>
<td>0.26*</td>
<td>0.27*</td>
<td>0.26*</td>
</tr>
<tr>
<td></td>
<td>CI: −0.35 to 0.60</td>
<td>0.15 to 0.31</td>
<td>−0.02 to 0.39</td>
<td>0.17 to 0.34</td>
<td>0.07 to 0.47</td>
<td>0.16 to 0.36</td>
</tr>
<tr>
<td></td>
<td>N: 7</td>
<td>26</td>
<td>16</td>
<td>43</td>
<td>15</td>
<td>22</td>
</tr>
</tbody>
</table>

*p ≤ .05.

School Staff Can Conduct Successful SEL Programs

Table 2 presents the mean effects obtained for the three major formats and supports the second hypothesis that school staff can conduct successful SEL programs. Classroom by Teacher programs were effective in all six outcome categories, and Multicomponent programs (also conducted by school staff) were effective in four outcome categories. In contrast, classroom programs delivered by nonschool personnel produced only three significant outcomes (i.e., improved SEL skills and prosocial attitudes, and reduced conduct problems). Student academic performance significantly improved only when school personnel conducted the intervention.

The prediction that multicomponent programs would be more effective than single-component programs was not supported (see Table 2). Multicomponent program effects were comparable to but not significantly higher than those obtained in Classroom by Teacher programs in four outcome areas (i.e., attitudes, conduct problems, emotional distress, and academic performance). They did not yield significant effects for SEL skills or positive social behavior, whereas Class by Teacher programs did.

What Moderates Program Outcomes?

We predicted that the use of the four SAFE practices to develop student skills and reported
implementation problems would moderate program outcomes, and in separate analyses we divided the total group of studies according to these variables. Both hypotheses regarding program moderators received support, and in separate analyses we divided the total group of studies according to these variables. Both hypotheses regarding program moderators received support, and the resulting mean ESs are presented in Table 3. Programs following all four recommended training procedures (i.e., coded as SAFE) produced significant effects for all six outcomes, whereas programs not coded as SAFE achieved significant effects in only three areas (i.e., attitudes, conduct problems, and academic performance). Reported implementation problems also moderated outcomes. Whereas programs that encountered implementation problems achieved significant effects in only two outcome categories (i.e., attitudes and conduct problems), interventions without any apparent implementation problems yielded significant mean effects in all six categories.

$Q$ statistics and $I^2$ values related to moderation. Table 4 contains the values for $Q$ and $I^2$ when studies were divided to test the influence of our hypothesized moderators. We used $I^2$ to complement the $Q$ statistic because the latter has low power when the number of studies is small and conversely may yield statistically significant findings when there are a large number of studies even though the amount of heterogeneity might be low (Higgins et al., 2003). To support moderation, $I^2$ values should reflect low within-group but high between-group heterogeneity. This would suggest that the chosen variable creates subgroups of studies each drawn from a common population, and that there are important differences in ESs between groups beyond what would be expected based on sampling error. $I^2$ values range from 0% to 100%, and based on the results of many meta-analyses, values around 15% reflect a mild degree of heterogeneity, between 25% and 50% a moderate degree, and values $\geq 75\%$ a high degree of heterogeneity (Higgins et al., 2003).

The data in Table 4 support the notion that both SAFE and implementation problems moderate SEL outcomes. For example, based on $I^2$ values, initially dividing ESs according to the six outcomes does produce the preferred low overall degree of within-group heterogeneity (15%) and high between-group heterogeneity (88%); for two specific outcomes, however, there is a mild (positive social behaviors, 32%) to moderately high (skills, 65%) degree of within-group heterogeneity. When the studies are further divided by SAFE practices or by implementation problems, the overall within-group variability remains low (12% and 13%, respectively), the

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Findings for Moderator Analyses at Post by Outcome Category for Total Sample</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Skills</th>
<th>Attitudes</th>
<th>Social behavior</th>
<th>Conduct problems</th>
<th>Emotional distress</th>
<th>Academic performance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Moderators</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommended training practices (SAFE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met SAFE criteria</td>
<td>ES 0.69*</td>
<td>0.24*</td>
<td>0.28*</td>
<td>0.24*</td>
<td>0.28*</td>
<td>0.28*</td>
</tr>
<tr>
<td>CI 0.52 to 0.86</td>
<td>0.18 to 0.29</td>
<td>0.18 to 0.38</td>
<td>0.18 to 0.31</td>
<td>0.14 to 0.42</td>
<td>0.17 to 0.38</td>
<td></td>
</tr>
<tr>
<td>N 63</td>
<td>80</td>
<td>73</td>
<td>88</td>
<td>33</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Did not meet SAFE criteria</td>
<td>ES 0.01</td>
<td>0.16*</td>
<td>0.02</td>
<td>0.16*</td>
<td>0.18</td>
<td>0.26*</td>
</tr>
<tr>
<td>CI −0.57 to 0.60</td>
<td>0.07 to 0.25</td>
<td>−0.21 to 0.26</td>
<td>0.04 to 0.28</td>
<td>−0.02 to 0.37</td>
<td>0.11 to 0.40</td>
<td></td>
</tr>
<tr>
<td>N 5</td>
<td>26</td>
<td>13</td>
<td>24</td>
<td>16</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Implementation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not mentioned</td>
<td>ES 0.58*</td>
<td>0.17*</td>
<td>0.32*</td>
<td>0.24*</td>
<td>0.21*</td>
<td>0.31*</td>
</tr>
<tr>
<td>CI 0.33 to 0.83</td>
<td>0.09 to 0.24</td>
<td>0.17 to 0.47</td>
<td>0.13 to 0.34</td>
<td>0.04 to 0.38</td>
<td>0.18 to 0.45</td>
<td></td>
</tr>
<tr>
<td>N 29</td>
<td>46</td>
<td>33</td>
<td>35</td>
<td>22</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>No problems</td>
<td>ES 0.86*</td>
<td>0.29*</td>
<td>0.31*</td>
<td>0.27*</td>
<td>0.35*</td>
<td>0.33*</td>
</tr>
<tr>
<td>CI 0.59 to 1.12</td>
<td>0.21 to 0.37</td>
<td>0.17 to 0.45</td>
<td>0.18 to 0.36</td>
<td>0.16 to 0.54</td>
<td>0.20 to 0.46</td>
<td></td>
</tr>
<tr>
<td>N 26</td>
<td>36</td>
<td>34</td>
<td>45</td>
<td>16</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Implementation problems</td>
<td>ES 0.35</td>
<td>0.19*</td>
<td>0.01</td>
<td>0.15*</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>CI −0.01 to 0.71</td>
<td>0.10 to 0.28</td>
<td>−0.18 to 0.19</td>
<td>0.05 to 0.25</td>
<td>−0.08 to 0.38</td>
<td>−0.01 to 0.28</td>
<td></td>
</tr>
<tr>
<td>N 13</td>
<td>24</td>
<td>19</td>
<td>32</td>
<td>11</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

Note. Means with subscript a differ significantly from each other at the .05 level. $^*p \leq .05$. 
within-group heterogeneity for both skills and social behaviors is no longer significant according to \( Q \) statistics, \( I^2 \) values drop to low levels (\( \leq 15\% \)) and remain low for the other outcomes as well, and heterogeneity levels attributed to differences between groups are high or moderate (\( I^2 \) values of 79\% and 63\% for SAFE and implementation, respectively). In other words, the use of all four SAFE practices and reported implementation problems to subdivide groups provided a good fit for the obtained data.

These latter findings are consistent with the mean differences between groups on many outcomes for the SAFE and implementation data presented in Table 3. SAFE and implementation problems were not significantly correlated (\( r = -0.07 \)). However, it was not possible to explore their potential interactions as moderators because only 57\% of the studies monitored implementation and subdividing the studies created extremely small cell sizes that would not support reliable results.

Inspection of the distribution of the moderator variables in the different cells in Table 3 indicated that SAFE practices and implementation problems were more common for some intervention formats. Compared to teacher-led programs, multicomponent programs were less likely to contain features that were significantly associated with better results for most outcomes, and may explain why the hypothesized superiority of multicomponent programs was not confirmed.

### Ruling Out Rival Hypotheses

After our primary analyses were conducted (see Table 2), we examined other possible explanations for these results. Additional analyses were conducted by collapsing across the three intervention formats and analyzing effects for the six outcome categories at post. First, we separately analyzed the impact of six methodological features (i.e., use of randomized designs, total and differential attrition, use of a reliable or valid outcome measure, and source of data: students vs. all others). We also analyzed outcomes as a function of students’ mean age, the duration of intervention (in both weeks and number of sessions), and the school’s geographical location (i.e., urban, suburban, or rural). We compared ESs for the three largest cells containing ethnicity data (Caucasian, \( k = 48 \); African American, \( k = 19 \); and Mixed, \( k = 75 \)). We also examined whether published reports yielded higher ESs than unpublished reports. Finally, we assessed if the three major intervention formats differed on any of the above variables (in addition to SAFE criteria and implementation problems) that might

---

### Table 4

\textbf{Q Statistics and }\textit{I^2} \textbf{ Values (in Percent) for Study Groupings for Moderator Analyses}

<table>
<thead>
<tr>
<th>Grouping variable</th>
<th>Values across all outcomes</th>
<th>Values within each outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( Q ) Between</td>
<td>( Q ) Within</td>
</tr>
<tr>
<td>All six outcomes</td>
<td>41.6*</td>
<td>530.2*</td>
</tr>
<tr>
<td>For each outcome</td>
<td>( Q ) within</td>
<td>193.9*</td>
</tr>
<tr>
<td></td>
<td>( I^2 ) within</td>
<td>65</td>
</tr>
<tr>
<td>SAFE practices</td>
<td>4.8*</td>
<td>74.8</td>
</tr>
<tr>
<td>For each outcome</td>
<td>( Q ) within</td>
<td>74.8</td>
</tr>
<tr>
<td></td>
<td>( I^2 ) within</td>
<td>12</td>
</tr>
<tr>
<td>Implementation</td>
<td>5.3*</td>
<td>75.0</td>
</tr>
<tr>
<td>For each outcome</td>
<td>( Q ) within</td>
<td>75.0</td>
</tr>
<tr>
<td></td>
<td>( I^2 ) within</td>
<td>13</td>
</tr>
</tbody>
</table>

*\( p \leq 0.05 \).
suggest the need for additional data analysis, but this latter procedure did not reveal any major differences across formats.

Findings. Among the 72 additional analyses we conducted (12 variables crossed with six outcomes) there were only four significant results, a number expected based on chance. Among the methodological variables the only significant finding was that for positive social behavior: Outcome data from other sources yielded significantly higher effects than those from student self-reports. The other three significant findings were all related to the skill outcome category. Students’ mean age and program duration were significantly and negatively related to skill outcomes ($rs = -0.27$ and $-0.25$), and published studies yielded significantly higher mean ESs for skills than unpublished reports. We also looked for potential differences within each of our outcome categories for ESs that were and were not adjusted for preintervention differences. The patterns of our major findings were similar (i.e., on such variables as teacher-effectiveness, use of SAFE practices, and implementation).

Effect of nested designs. In addition, all of the reviewed studies employed nested group designs in that the interventions occurred in classrooms or throughout the school. In such cases, individual student data are not independent. Although nested designs do not affect the magnitude of ESs, the possibility of Type I error is increased. Because few authors employed proper statistical procedures to account for this nesting or clustering of data, we reanalyzed the outcome data in Table 2 for all statistically significant findings following recommendations of the Institute of Education Sciences (2008a). These reanalyses changed only 1 of the 24 findings in Table 2. The mean effect for Class by Nonschool Personnel (0.17) was no longer statistically significant for conduct problems.

Possible publication bias. Finally, we used the trim and fill method (Duval & Tweedie, 2000) to check for the possibility of publication bias. Because the existence of heterogeneity can lead the trim and fill method to underestimate the true population effect (Peters, Sutton, Jones, Abrams, & Rushton, 2007), we focused our analyses on the homogeneous cells contained in Table 3 (e.g., the 112, 49, and 35 interventions with outcome data on conduct problems, emotional distress, and academic performance, respectively, and so on). The trim and fill analyses resulted in only slight reductions in the estimated mean effects with only one exception (skill outcomes for SAFE programs: original mean = 0.69; trim and fill estimate = 0.45). However, all the estimated means from the trim and fill analysis remained significantly different from zero. In sum, the results of additional analyses did not identify other variables that might serve as an alternative explanation for the current results.

Interpreting Obtained ESs in Context

Aside from SEL skills (mean ES = 0.57), the other mean ESs in Table 2 might seem ‘small.’ However, methodologists now stress that instead of reflexively applying Cohen’s (1988) conventions concerning the magnitude of obtained effects, findings should be interpreted in the context of prior research and in terms of their practical value (Durlak, 2009; Hill, Bloom, Black, & Lipsey, 2007). Table 5 presents the overall mean ESs obtained in the current review along with those obtained on similar outcomes from other meta-analyses of psychosocial or educational interventions for school-age youth, including several school-based prevention meta-analyses. Inspection of Table 5 indicates that SEL programs yield results that are similar to or, in some cases, higher than those achieved by other types of universal interventions in each outcome category. In particular, the postmean ES for academic achievement tests (0.27) is comparable to the results of 76 meta-analyses of strictly educational interventions (Hill et al., 2007).

It is also possible to use Cohen’s $U_3$ index to translate the mean ES on measures of academic

---

Table 5

Comparing Current Effect Sizes to Previous Meta-Analytic Findings for School-Age Populations

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Mean posteffects</th>
<th>Current review</th>
<th>Other reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skills</td>
<td>0.57</td>
<td>0.40&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Attitudes</td>
<td>0.23</td>
<td>0.09&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Positive social behaviors</td>
<td>0.24</td>
<td>0.39&lt;sup&gt;c&lt;/sup&gt;, 0.37&lt;sup&gt;d&lt;/sup&gt;, 0.15&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Conduct problems</td>
<td>0.22</td>
<td>0.26&lt;sup&gt;e&lt;/sup&gt;, 0.28&lt;sup&gt;e&lt;/sup&gt;, 0.21&lt;sup&gt;d&lt;/sup&gt;, 0.17&lt;sup&gt;c&lt;/sup&gt;, 0.30&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Emotional distress</td>
<td>0.24</td>
<td>0.21&lt;sup&gt;b&lt;/sup&gt;, 0.24&lt;sup&gt;c&lt;/sup&gt;, 0.17&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Academic performance</td>
<td>0.27</td>
<td>0.29&lt;sup&gt;b&lt;/sup&gt;, 0.11&lt;sup&gt;d&lt;/sup&gt;, 0.30&lt;sup&gt;f&lt;/sup&gt;, 0.24&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

Note. Results from other meta-analyses are from outcome categories most comparable to those in the current review, and values are drawn from weighted random effects analyses whenever possible.

performance into a percentile rank for the average student in the intervention group compared to the average control student who, by definition, ranks at the 50th percentile (Institute of Education Sciences, 2008b). A mean ES of 0.27 translates into a percentile difference of 11%. In other words, the average member of the control group would demonstrate an 11-percentile gain in achievement if they had participated in an SEL program. While higher ESs in each outcome area would be even more desirable, in comparison to the results of previous research, current findings suggest that SEL programs are associated with gains across several important attitudinal, behavioral, and academic domains that are comparable to those of other interventions for youth.

**Discussion**

Current findings document that SEL programs yielded significant positive effects on targeted social-emotional competencies and attitudes about self, others, and school. They also enhanced students’ behavioral adjustment in the form of increased prosocial behaviors and reduced conduct and internalizing problems, and improved academic performance on achievement tests and grades. While gains in these areas were reduced in magnitude during follow-up assessments and only a small percentage of studies collected follow-up information, effects nevertheless remained statistically significant for a minimum of 6 months after the intervention. Collectively, these results build on positive results reported by other research teams that have conducted related reviews examining the promotion of youth development or the prevention of negative behaviors (Catalano et al., 2002; Greenberg et al., 2001; Hahn et al., 2007; Wilson & Lipsey, 2007; Wilson et al., 2001).

The current meta-analysis differs in emphasis from previous research syntheses by focusing exclusively on universal school-based social-emotional development programs and evaluating their impact on positive social behavior, problem behaviors, and academic performance. Not surprisingly, the largest ES occurred for social-emotional skill performance (mean ES = 0.69). This category included assessments of social-cognitive and affective competencies that SEL programs targeted such as emotions recognition, stress-management, empathy, problem-solving, or decision-making skills. While it would be theoretically interesting to examine the impact of teaching various social versus emotional skills, SEL program designers typically combine rather than separate the teaching of these skills because they are interested in promoting the integration of emotion, cognition, communication, and behavior (Crick & Dodge, 1994; Lemerise & Arsenio, 2000). Thus, attempts to foster discrete emotions skills without also teaching social-interaction skills could be shortsighted from an intervention standpoint. However, for research and theoretical purposes, research designs that examine the relative contribution of different intervention components can help to determine which specific skills or combinations of skills lead to different outcomes at different developmental periods (Collins, Murphy, Nair, & Strecher, 2005).

Another important finding of the current meta-analysis is that classroom teachers and other school staff effectively conducted SEL programs. This result suggests that these interventions can be incorporated into routine educational practices and do not require outside personnel for their effective delivery. It also appears that SEL programs are successful at all educational levels (elementary, middle, and high school) and in urban, suburban, and rural schools, although they have been studied least often in high schools and in rural areas.

Although based on a small subset of all reviewed studies, the 11-percentile gain in academic performance achieved in these programs is noteworthy, especially for educational policy and practice. Results from this review add to a growing body of research indicating that SEL programming enhances students’ connection to school, classroom behavior, and academic achievement (Zins et al., 2004). Educators who are pressured by the No Child Left Behind legislation to improve the academic performance of their students might welcome programs that could boost achievement by 11 percentile points.

There are a variety of reasons that SEL programming might enhance students’ academic performance. Many correlational and longitudinal studies have documented connections between social-emotional variables and academic performance (e.g., Caprara, Barbaranelli, Pastorelli, Bandura, & Zimbardo, 2000; Wang et al., 1997). Compelling conceptual rationales based on empirical findings have also been offered to link SEL competencies to improved school attitudes and performance (Zins et al., 2004). For example, students who are more self-aware and confident about their learning capacities try harder and persist in the face of challenges (Aronson, 2002). Students who set high academic goals, have self-discipline, motivate
themselves, manage their stress, and organize their approach to work learn more and get better grades (Duckworth & Seligman, 2005; Elliot & Dweck, 2005). Also, students who use problem-solving skills to overcome obstacles and make responsible decisions about studying and completing homework do better academically (Zins & Elias, 2006). Further, new research suggests that SEL programs may affect central executive cognitive functions, such as inhibitory control, planning, and set shifting that are the result of building greater cognitive-affect regulation in prefrontal areas of the cortex (Greenberg, 2006).

In addition to person-centered explanations of behavior change, researchers have highlighted how interpersonal, instructional, and environmental supports produce better school performance through the following means: (a) peer and adult norms that convey high expectations and support for academic success, (b) caring teacher–student relationships that foster commitment and bonding to school, (c) engaging teaching approaches such as proactive classroom management and cooperative learning, and (d) safe and orderly environments that encourage and reinforce positive classroom behavior (e.g., Blum & Libbey, 2004; Hamre & Pianta, 2006; Hawkins et al., 2004; Jennings & Greenberg, 2009). It is likely that some combination of improvements in student social-emotional competence, the school environment, teacher practices and expectations, and student–teacher relationships contribute to students’ immediate and long-term behavior change (Catalano et al., 2002; Schaps et al., 2004).

As predicted, two variables moderated positive student outcomes: SAFE practices and implementation problems, suggesting that beneficial programs must be both well designed and well conducted. In the former case, current data replicate similar findings regarding the value of SAFE practices in after-school programs. In that review, programs that followed the same SAFE procedures were effective in multiple outcome areas, whereas those that failed to do so were not successful in any area (Durlak et al., 2010). Moreover, these findings are consistent with several other reviews that conclude that more successful youth programs are interactive in nature, use coaching and role playing, and employ a set of structured activities to guide youth toward achievement of specific goals (DuBois, Holloway, Valentine, & Cooper, 2002; Tobler et al., 2000).

Developing an evidence-based intervention is an essential but insufficient condition for success; the program must also be well executed. Although many studies did not provide details on the different types of implementation problems that occurred or what conditions were in place to ensure better implementation, our findings confirm the negative influence of implementation problems on program outcomes that has been reported in meta-analyses of other youth programs (DuBois et al., 2002; Smith, Schneider, Smith, & Ananiadou, 2004; Tobler et al., 2000; Wilson, Lipsey, & Derzon, 2003).

Contrary to our hypothesis, we did not find the expected additional benefit of multicomponent programs over single-component (i.e., classroom-only) programs, a finding that has been reported in other reviews of prevention and youth development interventions (Catalano et al., 2002; Greenberg et al., 2001; Tobler et al., 2000). In the current meta-analysis, this may be due to the fact that compared to classroom-only programs, multicomponent programs were less likely to follow SAFE procedures when promoting student skills and were more likely to encounter implementation problems. It is probable that the presence of one or both of these variables reduced program impact for many multicomponent interventions. For example, many multicomponent programs involved either or both a parent and schoolwide component, and these additional elements require careful planning and integration. Others have found that more complicated and extensive programs are likely to encounter problems in implementation (Durlak & Dupre, 2008; Wilson & Lipsey, 2007; Wilson et al., 2003). It is also important to point out that few studies compared directly the effects of classroom-based programming with classroom programming plus coordinated schoolwide and parent components (e.g., Flay, Graumlich, Segawa, Burns, & Holliday, 2004). An important priority for future research is to determine through randomized trials the extent to which additional components add value to classroom training.

How much confidence can be placed in the current findings? Our general approach and analytic strategy had several strengths: the careful search for relevant published and unpublished studies, testing of a priori hypotheses, and subsequent analyses ruling out plausible alternative explanations for the findings. We also reanalyzed our initial findings to account for nested designs that could inflate Type I error rates. Furthermore, we used only school records of grades and standardized achievement test scores as measures of academic performance, not students’ self-reports, and when examining follow-up results, we required data...
collection to be at least 6 months postintervention. Overall, findings from the current meta-analysis point to the benefits of SEL programming. Nevertheless, current findings are not definitive. The longitudinal research of Duncan et al. (2007) presented an alternative perspective in pointing out that attention skills, but not social skills, predict achievement outcomes. They noted, however, that social-emotional competencies may predict other mediators of school success such as self-concept, school adjustment, school engagement, motivation for learning, and relationships with peers and teachers. Future research on SEL programming can be improved in several ways to shed light on if and how newly developed SEL skills in school children relate to their subsequent adjustment and academic performance.

Limitations and Future Research Directions

More data across multiple outcome areas are needed. Only 16% of the studies collected information on academic achievement at post, and more follow-up investigations are needed to confirm the durability of program impact. Although all reviewed studies targeted the development of social and emotional skills in one way or another, only 32% assessed skills as an outcome. This is essential to confirm that the program was successful at achieving one of its core proximal objectives. Because there is no standardized approach in measuring social and emotional skills, there is a need for theory-driven research that not only aids in the accurate assessment of various skills but also identifies how different skills are related (Dirks, Treat, & Weersing, 2007). More rigorous research on the presumed mediational role of SEL skill development is also warranted. Only a few studies tested and found a temporal relation between skill enhancement and other positive outcomes (e.g., Ngwe, Liu, Flay, Segawa, & Aban-aya Co-Investigators, 2004). In addition, conducting subgroup analyses can determine if certain participant characteristics are related to differential program benefits. For example, factors such as ethnicity, developmental level, socioeconomic status, or gender may each influence who receives more or less benefit from an intervention (Reid, Eddy, Fetrow, & Stoolmiller, 1999; Taylor, Liang, Tracy, Williams, & Seigle, 2002; Wilson & Lipsey, 2007).

In addition to person-centered explanations for why SEL programming promotes positive outcomes, our findings indicate that it is important to attend to systemic and environmental factors (Greenberg et al., 2003). Programs that occur in classrooms or throughout the school are likely to be impacted by the organizational and ecological features of these environments. A few prevention and promotion studies have begun to explore the importance of classroom, school, and neighborhood context on program outcomes to illustrate how a broader ecological perspective can enhance our understanding of program effects (Aber, Jones, Brown, Chaudry, & Samples, 1998; Boxer, Guerra, Huesmann, & Morales, 2005; Metropolitan Area Child Study Research Group, 2002; Tolan et al., 1995). As a final example, analyses of the effects of the Child Development Project have indicated that improvements in the psychosocial environment of the school that were obtained during intervention mediated almost all of the positive student outcomes (Solomon, Battistich, Watson, Schaps, & Lewis, 2000).

More attention should focus on other potential moderators of program outcomes. We evaluated the composite effects of following four recommended practices (Sequential, Active, Focused, and Explicit) relating to effective skill training because previous authors have emphasized that these factors act in combination to produce better results. However, it is possible that some practices may be more important than others depending on the nature and number of targeted skills and the developmental abilities of students. For example, younger students may need more time to acquire more complex skills. Moreover, the four practices we evaluated do not capture every aspect of effective skill development such as procedures to encourage generalization of newly learned skills and training that is developmentally and culturally appropriate (Dusenbury & Falco, 1995; Gresham, 1995). We could not examine these other features due to lack of information in study reports, but their impact on skill development merits future attention. Furthermore, it would be preferable to evaluate SAFE practices as continuous rather than dichotomous variables. That is, program staff can be compared in terms of how much they focus on skill development and the extent of their use of active learning techniques instead of viewing these practices as all-or-none phenomena. An observational system has been developed to assess the use of SAFE practices as continuous variables in youth settings (Pechman, Russell, & Birmingham, 2008).

Although current results support the impact of implementation on outcomes, 43% of the studies did not monitor implementation in any way and thus were excluded from that analysis. Assessing
implementation should be seen as a fundamental and necessary aspect of any future program evaluations and efforts should be undertaken to evaluate the multiple ecological factors that can hinder or promote effective delivery of new programs (Durlak & Dupre, 2008; Greenhalgh et al., 2005).

**Raising Healthy Children: Implications for Policy and Practice**

Overall, research on school-based mental health and competence promotion has advanced greatly during the past 15 years. The Institute of Medicine’s (1994) first report on prevention concluded there was not enough evidence to consider mental health promotion as a preventive intervention. However, the new Institute of Medicine (2009) report on prevention represents a major shift in thinking about promotion efforts. Based on its examination of recent outcome studies, the new Institute of Medicine report indicated that the promotion of competence, self-esteem, mastery, and social inclusion can serve as a foundation for both prevention and treatment of mental, emotional, and behavioral disorders. The Report of the Surgeon General’s Conference on Children’s Mental Health expressed similar sentiments about the importance of mental health promotion and SEL for optimal child development and school performance by proclaiming: “Mental health is a critical component of children’s learning and general health. Fostering social and emotional health in children as a part of healthy child development must therefore be a national priority” (U.S. Public Health Service, 2000, p. 3).

Although more research is needed to advance our understanding of the impacts of SEL programming, it is also important to consider next steps for policy and practice at the federal, state, and local levels. At the federal level, there is bipartisan sponsorship of HR 4223: The Academic, Social, and Emotional Learning Act. This bill authorizes the Secretary of Education to award a 5-year grant to establish a National Technical Assistance and Training Center for Social and Emotional Learning that provides technical assistance and training to states, local educational agencies, and community-based organizations to identify, promote, and support evidence-based SEL standards and programming in elementary and secondary schools. A recent review of U.S. school practices found that 59% of schools already have in place programming to address the development and support of children’s social and emotional competencies (Foster et al., 2005). It is critical to ensure that these efforts are informed by theory and research about best SEL practice. Incorporating provisions of HR 4223 into the reauthorization of the Elementary and Secondary Education Act will help to achieve that objective.

Furthermore, there are active efforts in some states (e.g., Illinois, New York) and internationally (e.g., Singapore) to establish and implement SEL standards for what students should know and be able to do. For example, as the result of recent legislative action, Illinois became the first state to require every school district to develop a plan for the implementation of SEL programming in their schools. In addition, the Illinois State Board of Education recently incorporated SEL skills as part of their student learning standards, identifying three broad learning goals: (a) develop self-awareness and self-management skills to achieve school and life success, (b) use social awareness and interpersonal skills to establish and maintain positive relationships, and (c) demonstrate decision-making skills and responsible behaviors in personal, school, and community contexts (see http://isbe.net/ils/social_emotional/standards.htm). Increasingly, policymakers at the federal, state, and local level are embracing a vision of schooling in which SEL competencies are important.

Unfortunately, surveys indicate that many schools do not use evidence-based prevention programs or use them with poor fidelity (Gottfredson & Gottfredson, 2002; Ringwalt et al., 2009). This may occur for a variety of reasons: Schools may not be aware of effective programs, fail to choose them from among alternatives, do not implement the interventions correctly, or do not continue programs even if they are successful during a pilot or demonstration period. In other words, there is a wide gap between research and practice in school-based prevention and promotion just as there is with many clinical interventions for children and adolescents (Weisz, Sandler, Durlak, & Anton, 2005).

If effective programs are to be used more widely, then concerted efforts are needed to help schools through the multiple steps of the diffusion process. These steps include the dissemination of information about available programs, adoption of programs that fit best with local settings, proper implementation of newly adopted programs, effective program evaluation to assess progress toward desired goals, and methods to sustain beneficial interventions over the long term (Wandersman & Florin, 2003). A variety of efforts are needed
to develop state and local capacity to encourage widespread evidence-based programming (Fixsen, Naoom, Blase´, Friedman, & Wallace, 2005). It is especially important to document the costs and benefits of prevention programming. Recent analyses suggest that some SEL programs (e.g., Hawkins et al., 2004) are a good financial investment; however, future studies must include more cost analyses in their evaluation designs (Aos, Lieb, Mayfield, Miller, & Pennucci, 2004). With adequate funding, capacity can be built through providing policy supports, professional development, and technical assistance to promote educator knowledge and motivation for the best ways to identify, select, plan, implement, evaluate, and sustain effective SEL interventions (Devaney, O’Brien, Resnik, Keister, & Weissberg, 2006; Osber, Dwyer, & Jackson, 2004). Effective leadership and planning also promote quality program implementation through ensuring adequate financial, personnel, and administrative support as well as providing professional development and technical assistance (Devaney et al., 2006; Kam, Greenberg, & Walls, 2003). Along with this effective planning and programming, there is a need to establish assessment and accountability systems for SEL programs in relation to student outcomes (Greenberg et al., 2003; Marzano, 2006). Addressing these issues will increase the likelihood that more evidence-based programs will be effectively implemented and sustained in more schools, which, in turn, will support the healthy academic, social, and emotional development of more children.

References

References marked with an asterisk indicate studies included in the meta-analysis.


Cunningham, E. G., Brandon, C. M., & Frydenberg, E. (2002). Enhancing coping resources in early adolescence through a school-based program teaching optimistic thinking skills. Anxiety, Stress, and Coping, 15, 369–381.


Abstracts International: Section B. Sciences and Engineering, 58(12), 6803.


*Heinemann, G. H. (1990). The effects of the Lions-Quest “Skills for Adolescence” program on self-esteem development and academic achievement at the middle school level. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 51*(6), 1890A.


*Hunter, L. (1998). Preventing violence through the promotion of social competence and positive interethnic
contact: An evaluation of three elementary school-based violence prevention instructional approaches. Dissertation Abstracts International: Section A. Humanities and Social Sciences, 59(08), 2851.


*LeMing, J. S. (2001). Integrating a structured ethical reflection curriculum into high school community


*Jr., M. C., Wang, & H. J. Walberg (Eds.), *Building academic success on social and emotional learning: What does the research say? (pp. 189–205). New York: Teachers College Press.


