Effective Components of Social Emotional Learning Programs: A Meta-analysis

Jieping Shi [1] · Alan C. K. Cheung[2]

Abstract

Previous reviews have synthesized the impacts of universal school-based social emotional learning (SEL) programs. However, they have yet to attempt a meta-analytic approach with rigorous inclusion criteria to identify the key SEL components and explore what make these programs work. This study aims to fill that gap by examining the impacts of SEL programs and exploring the moderating effects of methodological characteristics, implementation features, and program components on SEL effectiveness. The final sample consisted of 12 high-quality SEL programs, 59 studies, and 83,233 participants, with an overall effect size of 0.15. Meta-regression results indicated that these SEL programs could significantly improve youth social emotional skills, reinforce affect and attitudes, promote academic performance, increase prosocial behaviors, and reduce antisocial behaviors. Training teachers' social emotional skills and reducing cognitive elements in SEL curricula were found to be effective components of SEL programs, whereas pedagogical activities, climate support, and family engagement were not. Large-scale studies of SEL programs tended to generate smaller effect sizes, and those with low program dosages were found to be less effective than those approaching the recommended dosage. Policy and practical implications on how to scale SEL programs are discussed.

Introduction

There is general consensus that social emotional skills play a crucial role in personal success in the labor market and in behavioral outcomes (e.g., prosocial behaviors) and are at least as important as cognitive skills (Heckman et al., 2006). How to cultivate students' social emotional skills in schools has drawn increasing attention from teachers, parents, policy makers and other stakeholders over the years (Miyamoto et al., 2015). In the past two decades, school-based social emotional learning (SEL) programs have been widely adopted in various school settings across the world. SEL refers to the process by which children "acquire and apply the knowledge, skills, and attitudes to develop healthy identities, manage emotions and achieve personal and collective goals, feel and show empathy for others, establish and maintain supportive relationships, and make

responsible and caring decisions", which is proposed by the leading organization, Collaborative for Academic Social Emotional Learning (CASEL; 2013, p. 4). Numeral empirical research has indicated that SEL programs generally produce positive results on multiple outcomes for students (Low et al., 2019). Over a dozen meta-analyses on universal SEL programs have also confirmed similar positive effects, including but not limited to improving social emotional skills, attitude towards self and others, prosocial behaviors, academic performance, and reducing conduct problems, emotional distress, and substance abuse (Durlak et al., 2011). Though many experimental studies and research reviews on SEL programs have been carried out in the past few decades, none has attempted to use a meta-analysis approach to identify the key components (e.g., course content, pedagogical activities, teacher training) that make these programs work. Therefore, this study attempts to conduct a comprehensive meta-analysis to identify the program components that significantly predict SEL effectiveness. It will also examine the moderating effects of methodological features (e.g., research design, sample size) and implementation features (e.g., dosage) and compare the predictive effects of all moderators through meta-regression

✉ Jieping Shi
   shijieping@bnu.edu.cn

1   Faculty of Education, Beijing Normal University, Beijing, China

2   Department of Educational Administration & Policy, The Chinese University of Hong Kong, Hong Kong, China

models. Since this review categorizes the outcomes into several domains, it will provide a vivid picture of the effectiveness of SEL programs on different outcome domains.

Previous Research of SEL Programs

A set of reviews has been conducted in the area of SEL programs, which provides valuable insights and suggestive evidence regarding their effectiveness. However, these reviews have certain limitations, such as the qualitative nature, methodological constraints, or limited scope. None of these reviews specifically answer the key research question, i.e. to identify the core program components and explore their relations with the program effects.

Meta-analytic Research of SEL Programs

Perhaps the most common type of research review on SEL is the meta-analytic research, which included all SEL programs. This type of research generally provided an overview of SEL effect sizes in multiple outcome domains and identified some moderators that were mainly methodological. For instance, one widely-cited review summarized 213 universal school-based SEL programs, indicating that these programs significantly reinforced social emotional skills, academic achievements, attitudes and behaviors (Durlak et al., 2011). It examined the moderating effects of intervention formats, implementation features, and methodological variables, revealing that the employment of sequenced, active, focused, and explicit practices to develop skills (SAFE) and implementation problems significantly moderated SEL effects. However, these criteria were somewhat broad and inclusive for the high-quality SEL programs. For example, it only distinguished whether the included studies had implementation problems (such as not completing all the SEL requirements) without further categorizing what kind of the implementation problems mattered (Durlak et al., 2011). It could only indicate that SEL programs should be implemented as intended rather than providing other specific practical suggestions. The SAFE criteria were substantial quality markers that significantly moderated outcomes of all SEL interventions, including both efficacy and large-scale effectiveness trials. But dozens of high-quality and widely-used SEL programs, such as PATHS (Greenberg & Kusché, 2006), Second Step (Frey et al., 2000), Positive Action (Flay & Allred, 2003), 4Rs (Jones et al., 2011), RULER (Brackett et al., 2012), Open Circle (Hennessey, 2007), etc. all meet the SAFE criteria. These programs still have substantive differences in many aspects, including but not limited to theory foundation, course content, pedagogical approach, teacher training, family engagement, which may contribute to the differences in effectiveness.

Another review involving 89 SEL programs for students aged 4 to 18 years showed significant positive effects, with the largest effect on social emotional skills and the smallest effect on attitudes toward self (Wigelsworth et al., 2016). It further explored the moderation effects of evaluation stage, developer involvement, and implementation site, but the results were not always in the expected direction. For example, it hypothesized that higher developer involvement was associated with larger effect sizes, but their results did not support their hypothesis (Wigelsworth et al., 2016). These unexpected results might be due to the approach of subgroup analysis, which was unable to control for the interference of other moderators as multivariate meta-regression would. A third review involving 75 SEL programs also found positive effect sizes on multiple outcomes (Sklad et al., 2012). It extracted multiple moderators and explored their effects on social skills and antisocial behaviors respectively. Results showed that outcome source, manual, school level, duration, and number of sessions were significant moderators, while other factors, like location, trainers, and involved professionals, were not statistically significant (Sklad et al., 2012). All these three reviews employed very similar outcome domains, which could be referred to by the current review. Additionally, there were several other meta-analytic reviews on SEL (Table 1), including one investigating the follow-up effects of SEL interventions (Taylor et al., 2017), another examining the effects of SEL on academic performance (Corcoran et al., 2018), one focusing on updated SEL interventions from 2008 to 2020 (Cipriano et al., 2023), and four separate studies investigating the effects of preschool SEL interventions (Blewitt et al., 2018; Murano et al., 2020; Sabey et al., 2017; Yang et al., 2019).

There is a general consensus about the positive impacts of school-based SEL programs on students, with most effect sizes ranging from 0.09 to 0.70. These meta-analytic reviews substantially contributed to SEL research and practice, but they were not without limitations regarding the effects of moderators: (1) most extracted moderators were methodological variables, like research design, grade, duration, etc. Even though their moderation effects were partially consistent, these studies shed little light on how best to implement SEL intervention in school settings; (2) some reviews extracted implementation features as moderators, like SAFE criteria and developer involvement, but their findings were less suggestive and practical owing to inappropriate coding methods and statistical analysis approaches; (3) most reviews showed high heterogeneity levels, which might be due to the substantive differences of included SEL interventions in terms of course content, pedagogical approaches, teacher training, family engagement, etc., but none of them had attempted to identify such components and explore their moderation effects on SEL effectiveness. The current study builds firmly upon these

Table 1 The effect sizes of previous SEL meta-analytic reviews

Study	Student Grade	Number of studies	Effect sizes (Post)	Effect sizes (Follow up)	Moderators			
					Methodological	Implementation	Program components	Other
Durlak et al. (2011)	K-12	213	0.22–0.57	0.11–0.32	randomization, attrition, outcome measure, data source+, duration+, mean age+, location	implementation problems++	SAFE++, intervention format	publish status+
Sklad et al. (2012)	K-8	75	0.09–0.70	0.07–0.26	outcome source++, manual++, school level+, duration+, location	number of sessions+, trainers, professionals involved	focused skills	NA
Wigelsworth et al. (2016)	aged 4–18 years	89	0.17–0.53	NA	home/away+	efficacy/effectiveness+, developer involvement+	NA	NA
Cipriano et al. (2023)	K-12	424	0.061–0.293	0.116–0.267	duration+, home/away+, age+	delivery format++, implementation quality+	SAFE+, academic integration+	NA
Taylor et al. (2017)	K-12	82	NA	0.13–0.33	randomization, outcome measures, outcome source, length of follow-up, attrition+, duration, average age+, race, gender, sample size, location, population	number of sessions, implementation quality	SAFE, intervention format, tailored content	NA
Corcoran et al. (2018)	Pre-K-12	40	0.19–0.26	NA	research design, sample size+, grade level, socio-economic status	program intensity	NA	publish status, publication year
Sabey et al. (2017)	preschool	29	−0.08–0.38	NA	NA	NA	NA	NA
Blewitt et al. (2018)	preschool	63	0.18–0.54	NA	randomization, pretest, age of children, sex, socio-economic status, parent informant+, teacher informant, measures+, study quality	instruction time, length of program, intervention leader, mode of delivery	program's age, parental involvement	publication status
Yang et al. (2019)	preschool	15	0.15–0.24	NA	duration++, type of measures, quality of study design	implementation fidelity++	type of curriculum++	NA
Murano et al. (2020)	preschool	48	0.32–0.50	NA	randomization++, attrition, participant socio-economic status, race, risk status ++, outcome measures++	delivery format++, implementation fidelity++	parental involvement++	NA

++ means the factor could significantly moderate all SEL outcome categories, + means the factor could significantly moderate some of the SEL outcome categories, no + means the factor was not significant in any SEL outcome categories

meta-analytic reviews (e.g., Durlak et al., 2011) and explores the overall effects of SEL programs and the moderating roles of methodological variables, implementation features, and program components on SEL effectiveness.

Reviews of One Single SEL Program: Implementation Matters

Four previous reviews focused exclusively on one single SEL program. Since choosing one program largely decreases the heterogeneity of included studies, this type of review could focus more on the moderation effects of methodological and implementation features in-depth. One review examined the effects of the PATHS program on SEL skills, attitudes, behaviors, and academic performance, and found that dosage (i.e., the frequency of delivered lessons per week) was the most determinant factor of program effects among all moderators (Shi et al., 2022). The findings also indicated that dosage was more effective and distinguishable than other implementation features, like program quality and implementation fidelity (Shi et al., 2022). Another review of PATHS only summarized five studies in preschool settings and did not provide any valuable findings about moderators (Stanley, 2019). The last two reviews (i.e., Moy et al., 2018; Moy & Hazen, 2018) examined the overall effects of Second Step on knowledge, prosocial and antisocial outcomes and found that the program had much higher effects on antisocial outcomes than other outcomes. They further explored the moderation effects of five methodological and implementation factors on the three outcomes, showing that the effects of program saturation, outcome source, metro area, and location were not statistically significant (Moy et al., 2018).

These studies underscored the importance of SEL implementation, which was consistent with other SEL implementation research. One empirical study innovatively proposed nuanced measures of SEL implementation dosage and revealed that these dosage measures were positively associated with SEL outcomes like student attendance and adaptive behavior (Wu et al., 2023). Implementation research insisted that SEL outcomes could be more robust if the interventions were carefully implemented with adherence to the program requirements (Durlak & DuPre, 2008). Regarding the implementation information of SEL experimental studies, most studies provided information on implementation dosage, and only a few reported implementation quality or participant responsiveness (Berkel et al., 2011).

Guides of Selected SEL Programs

Unlike the two aforementioned reviews, this type of research provides a general overview of selected SEL programs, particularly looking into some of the widely used

SEL programs in the field to see how they differed from each other and what made each program unique (Jones & Bouffard, 2012). It viewed SEL not as a program but as a set of practices (Wigelsworth et al., 2022), since SEL programs employed an increasingly comprehensive and holistic approach. This type was more like a resource or program guide particularly for practitioners like teachers, principals, and school districts. It did not employ the meta-analysis approach which was commonly used in the first two types of reviews, prioritizing program descriptions and comparisons. CASEL developed a framework to evaluate the quality of current SEL programs, selected 23 programs for preschool and elementary schools (CASEL, 2013) and 12 programs for secondary schools (CASEL, 2015), and summarized their program features, which would be helpful for principals to select and implement SEL in their schools. Another guide selected 25 high-quality SEL programs and described their detailed features, including in-school and out-of-school programs (Jones et al., 2017). This document systematically summarized the skill focuses, instructional methods, and program components of selected high-quality SEL programs. Another study identified and compared the core components of 14 SEL programs and found the most common components were social skills and identifying others' feelings (Lawson et al., 2019). The last one identified core components of SEL programs by conducting a cross-tabulation analysis of common practice and instructional elements (i.e., what SEL is taught and how SEL is taught) (Wigelsworth et al., 2022). Unlike a meta-analysis approach, these reviews did look inside and compare the components of each program systematically and provided practical suggestions for stakeholders. Their comparison frameworks laid a solid foundation for identifying components in the current review. However, they failed to link the components to the outcomes, making it difficult to identify which components really worked in these comprehensive programs. While the previous studies focused on the components of SEL programs by emphasizing "what and how SEL is taught", the current study sought to explore "what works" by conducting a meta-analytic synthesis of SEL programs.

Potential Moderators on SEL Effectiveness

Despite limitations in previous research, they still underpinned the current study regarding the identification of potential moderators: (1) the classification of SEL outcomes and coding of methodological moderators from meta-analytic reviews; (2) the extraction and coding of potential implementation moderators from reviews of individual programs; (3) systematic comparisons of program components from the guides of selected SEL programs. It should be noted that there may be some overlaps between the

descriptions of previous SEL research and potential moderators, thus only a brief overview of relevant moderators (i.e., methodological variables, implementation features, and program components) is discussed here.

Multiple methodological factors, such as research design, sample size, grade level, and duration, may influence SEL outcomes to some extent. Randomized controlled trials have been found to produce smaller effect sizes compared to quasi-experimental studies (Cheung & Slavin, 2016), though this difference was sometimes insignificant (Taylor et al., 2017). Studies with smaller sample sizes also showed significantly higher effects than those with larger ones (Shi et al., 2022), but some meta-analytic reviews found that the sample size was an insignificant moderator (Corcoran et al., 2018) or did not analyze it as a moderator (Sklad et al., 2012). Grade level has shown to be a significant moderator of SEL outcomes in some studies (Moy et al., 2018), but not in others (Corcoran et al., 2018). Duration tended to be negatively associated with SEL outcomes (Yang et al., 2019), although this was not always the case (Blewitt et al., 2018). There was no consensus on the moderating effects of methodological variables on SEL outcomes. As to the implementation features, dosage was considered as one of the most critical and accessible factors. Most SEL experimental studies reported dosage information, while often neglecting other implementation features like quality (Berkel et al., 2011). Emerging research suggested that dosage was positively related to SEL outcomes (Wu et al., 2023). However, most previous meta-analytic reviews of SEL programs did not examine it as a moderator.

Based on previous SEL guides viewing SEL as a set of modules, the current review identified five key components, including cognitive elements, pedagogical activities, teacher social emotional skills, climate support, and family engagement. The skills targeted in the course content and pedagogical activities were undoubtedly the most critical components, which were systematically summarized and compared among different SEL programs (Wigelsworth et al., 2022). Almost all SEL programs placed a strong emphasis on social and emotional skills, while the proportion of cognitive content (e.g., working memory and cognitive flexibility) varied substantially from program to program (Jones et al., 2017). As such, cognitive elements in the course content seemed to be a core program component that should be considered. In coordination with program goals and focused skills, SEL programs generally involved various pedagogical activities, including but not limited to discussion, drawing, music, and games. There were apparent differences in the pedagogical approaches used among SEL programs (Jones et al., 2017). This variety or richness of pedagogical activities might be related to the effectiveness of SEL programs. Teacher social emotional skills, climate support, and family engagement were also identified

as potential moderators of SEL outcomes. For example, one study on the prosocial classroom model proposed that teacher SEL skills and supportive climate could contribute to positive social, emotional, and academic outcomes for youth and adolescents (Jennings & Greenberg, 2009). Family engagement has been linked to improved cognitive skills and social emotional skills by offering more opportunities for students to practice what was taught in schools (Kelty & Wakabayashi, 2020). This brief synthesis of potential methodological, implementation, and component moderators on SEL effectiveness provided the basis for the current review.

Current Study

Although previous SEL research has provided insights into the impact of SEL programs and the potential factors related to SEL effectiveness, none of them have identified the core program components and related methodological and implementation moderators that make programs effective. The current review is to conduct an updated review of the effects of SEL programs and figure out the most effective components through meta-analysis. It attempts to examine the impacts of SEL programs and to explore the moderating effects of the aforementioned methodological characteristics, implementation features, and program components on SEL effectiveness. Note that the current review has two premises. First, the SEL programs must be universal, high-quality, and curriculum-based programs. The corresponding reasons are as follows: (1) targeted SEL programs that focus on some particular types of students rather than all students, such as low socioeconomic status students and special education students may affect the effects; (2) high-quality SEL programs tend to be comprehensive and involve more components, and the number of their studies is relatively large, allowing for program comparison; (3) most high-quality SEL programs have scripted curricula for students, which have been examined to be effective. These criteria are similar to the previous studies on program comparison, perhaps with the exception of the curriculum-based aspect. Second, the current review employs a set of rigorous criteria for included studies to minimize the influence of irrelevant factors based on the best-evidence approach. After selecting universal, high-quality, and curriculum-based SEL programs, it identifies eligible studies of these programs in accordance with a set of stringent criteria (e.g., sufficient sample size, initial equivalence). The rigorous criteria for identifying included studies are aligned with the criteria for selecting SEL programs, ensuring the consistency and coordination of the current review. The research questions are threefold. First, what are the overall effects of these high-quality SEL programs? Second, what methodological characteristics, implementation features, and program components

are related to the effectiveness of these programs? Third, what are the most effective components of these programs?

Methods

Selection of SEL Programs

In order to examine the effective components of universal, high-quality and curriculum-based SEL programs, we need to first identify the eligible programs. The selection of SEL programs in the current review was mainly based on two SEL guides (CASEL, 2013; Jones et al., 2017). They identified 23 "SELect" programs and 25 leading programs respectively, and then compared the elements of these programs systematically. Based on these two fundamental and comprehensive reports, the current review adopted the common programs in these two studies to ensure the reliability of the selection process. The selection of programs was roughly consistent with one previous study on SEL components (Lawson et al., 2019). In sum, 12 eligible SEL programs were included in the current review, namely 4Rs, Caring School Community, Competent Kids Caring Communities, I Can Problem Solve, Lionzs Quest, MindUP, Open Circle, PATHS, Positive Action, RULER, Second Step, and Too Good For Violence.

According to the systematic comparative framework of the two SEL guides, we extracted five components of SEL programs. The operational definitions of these features were as follows: (1) cognitive elements, referring to whether a program involved a relatively high proportion of cognitive content (e.g., literacy) or learning skills (e.g., attention) in the course content; (2) pedagogical activities, referring to whether the program used various teaching and learning activities or pedagogical approaches for instruction (e.g., discussion, role-play, drawing, music, etc.); (3) teacher social emotional skills, referring to whether the program emphasized the cultivating of teacher social emotional assets when training teachers; (4) climate support, referring to whether the program aimed to reinforce a supportive climate in classrooms or schools (e.g., providing Principal Toolkit to reinforce positive culture in schools); (5) family engagement, referring to whether the program provided extensive family activities (e.g., parent-teacher conferences). The performance of these programs regarding each component is shown in Table S2 (see appendix 2) and is basically consistent with previous SEL guides (Jones et al., 2017). These components will be used as program moderators to examine their effectiveness.

Searching Procedures

The current review employed three searching strategies to find all potential studies. First, we used a set of searching strings to search in the academic databases, including Web of Science, Proquest, ERIC, and PsycINFO (see appendix 1 searching procedures). As for the programs with limited articles, we restricted the search to the names of programs in the Title or Abstract. As for the programs with many searching results, like Second Step, we used "impact or effect or effectiveness or evaluation or assessment" and "student or children or school" to refine the search results. Second, we searched the references in previous reviews on SEL. Third, we searched the website of each program to locate any unpublished grey studies or reports. These searching strategies ensured that we collected as many potential articles as possible.

Criteria for Included Studies

The current review employed rigorous inclusion criteria to select eligible studies of the twelve SEL programs. Adopting stringent inclusion criteria could reduce the interference caused by research quality in the analysis process, leading to more convincing conclusions. The inclusion criteria were as follows.

1. It must be written in English for practical reasons.
2. It must have been published between 1980 and 2020.
3. It must focus on the effects of the 12 high-quality SEL programs on students. Studies only involving the SEL impacts on teachers (e.g., teacher burnout, teacher well-being) or classrooms (e.g., classroom climate) rather than student outcomes were excluded (e.g., Sandilos et al., 2020).
4. The treatment group must implement only one single SEL program. If the treatment group used the combination of these SEL programs and other programs, the study would be excluded (e.g. Bradshaw et al., 2020).
5. It must have a control group. Studies without comparison groups were excluded (e.g., Wilson, 2016). And if the control group also involved SEL, the study would be excluded (e.g. Schonfeld et al., 2015).
6. Sample size must be at least 60 (30 for control group and 30 for treatment group), since experimental studies with small sample sizes tended to generate inflated effects (Cheung & Slavin, 2016). The current cutoff regarding sample size was consistent with some previous reviews (e.g., Neitzel et al., 2022). Studies with insufficient numbers of participants would be excluded for statistical reasons (e.g. Aras & Aslan, 2018).
7. The treatment and control group need to have initial equivalence at pretest(s) in terms of outcome measurements. The pretest differences on outcome

Fig. 1 PRISMA Flow chart of searching procedures
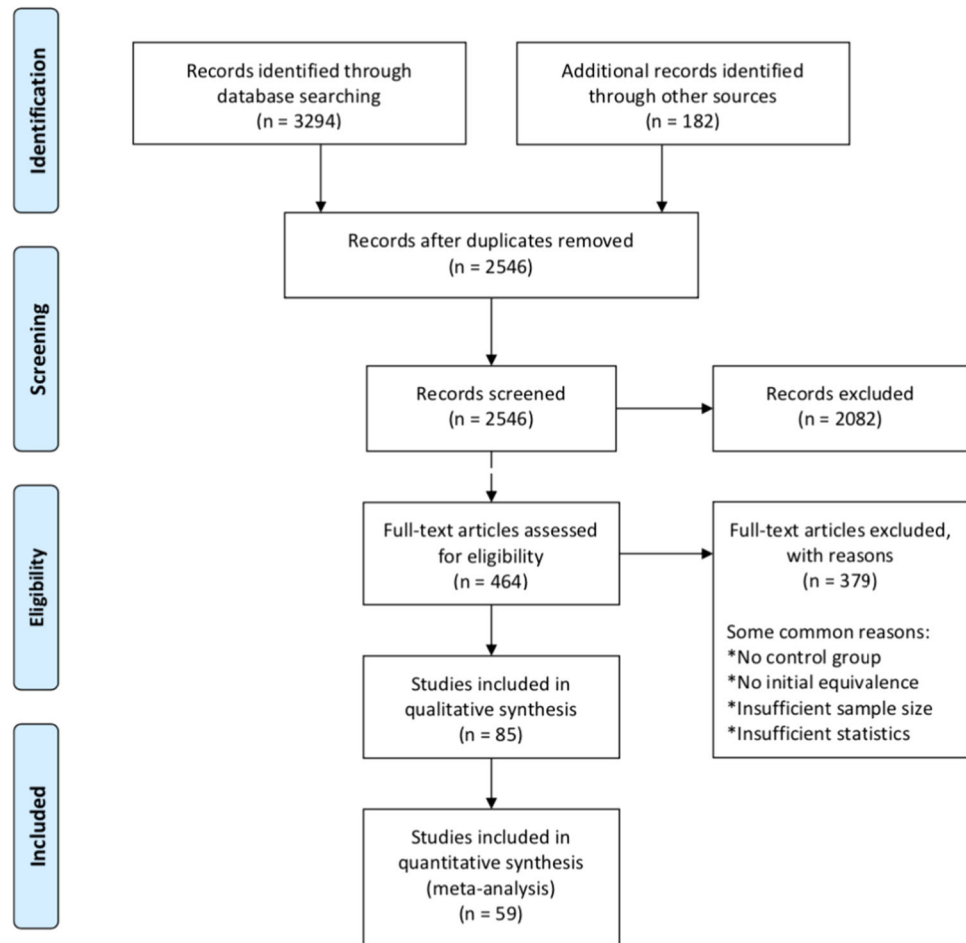
variables between treatment group and control group must be less than 0.25 standard deviation (What Works Clearinghouse, 2020). If a study had no initial equivalence or lacked relevant statistics, it would be excluded (e.g. Fishbein et al., 2016).

8. It must have enough quantitative statistics to estimate effect sizes at posttest(s). If a study only reported the effects at follow-up, it would be excluded (e.g. Averdijk et al., 2016).

A total of 3476 records were obtained after the initial search procedures. After removing 930 duplicates, only 464 studies of the 2546 records were regarded as potentially relevant in the title and abstract screening. In the process of full-text screening and coding, 379 records were excluded for reasons such as no control group, insufficient sample size, and no initial equivalence. Consequently, 85 articles met the inclusion criteria but only 59 studies were included because 26 articles used duplicated samples (e.g. Duncan et al., 2017). The flow chart exhibits the detailed searching procedures (Fig. 1).

Coding and Variables

As to the dependent variables, SEL outcomes were categorized into four domains, namely social emotional skills, affect and attitudes, behaviors and academic performance, adapted on the basis of previous outcome classifications (Durlak et al., 2011; Sklad et al., 2012; Wigelsworth et al., 2016). Note that the negative outcomes were transformed into a positive direction, so the positive effect sizes meant that SEL programs could improve positive outcomes and reduce negative outcomes.

(1) Social emotional skills. This category included not only self-awareness, self-management, social awareness, relationship skills, responsible decision-making (CASEL, 2013), but also relevant skills including but not limited to emotional intelligence, self-esteem, self-concept, self-efficacy, self-control, empathy, problem solving, conflict resolution, social skills, leadership. Measurements like Devereux Student Strengths Assessment (LeBuffe et al., 2009), Social Skills Rating System (Gresham & Elliott, 1990),

Rosenberg Self-Esteem Scale (Rosenberg, 1989), etc. were all included.

(2) Affect and attitudes. This domain mainly contained students' emotional distress, as well as their perceptions or attitudes towards peers, teachers and schools. For instance, anxiety, depression, stress, psychological well-being, school bonding, school orientation, peer relation, teacher relations, etc. were included in this domain. Measurements included Positive and Negative Affect Scale for Children (Laurent et al., 1999), Student Life Satisfaction Scale (Huebner, 1991), etc.

(3) Prosocial and antisocial behaviors. This category involved both prosocial behaviors and antisocial behaviors, including but not limited to altruistic behaviors, aggression, bullying, violence, disruptive behaviors, and substance abuse. Relevant measurements included Aggression Scale (Orpinas & Frankowski, 2001), Problem Behavior Frequency Scale (Farrell et al., 2016), etc.

(4) Academic performance. This domain was mainly obtained from school records, including but not limited to reading test scores, mathematics test scores, and academic competence.

In addition to the four dependent variables, several moderators that are worthy of investigation were identified. Methodological characteristics, implementation features, and program components were all coded as moderators or independent variables. Methodological characteristics included research design, sample size, grade level and duration, which had been confirmed to be related to the effects of SEL by previous reviews (Corcoran et al., 2018). Implementation features only had one factor, dosage, which had been examined to have a significant impact on program effects (Shi et al., 2022). Program components had five factors, including cognitive elements, pedagogical activities, teacher social emotional skills, climate support, and family engagement. Methodological and implementation variables were coded as follows: (1) research design, studies were coded as quasi-experiment or randomized controlled trial; (2) sample size, the numbers of participants in each study were coded as small ($N < 250$) or large ($N \geq 250$; Corcoran et al., 2018); (3) grade level, studies were coded as Pre-kindergarten or elementary or secondary; (4) duration, the studies were coded as one year or more than one year (Sklad et al., 2012); (5) dosage, the studies were coded as standard or low based on whether the number of delivered lessons was equivalent to a minimum of 80% coverage of program requirements (Shi et al., 2022). Two researchers coded all variables and effect sizes separately, and their coding results were highly consistent. The coding reliabilities of the two researchers were 95% on effect sizes and 100% on moderators. Any disagreements of coding were resolved through discussions.

Statistical Analysis

Regarding the effect size calculation for each included study, standardized mean difference Cohen's d (Cohen, 1987) was employed to estimate the effect size of included studies, quantifying the difference between the treatment and control groups divided by pooled standard deviations (Borenstein et al., 2021). If these statistics (i.e., sample size, means, and standard deviations) were not reported in the original studies, t-values or F-values were extracted between the treatment and control groups alternatively to calculate Cohen's d. Since the numbers of participants varied substantially among studies, Cohen's d was transferred to Hedges' g (Hedges, 1981) in the software of Comprehensive Meta-Analysis version 3 (Borenstein et al., 2013) to compensate for the impacts of the sample size. For the studies with multiple outcomes, all effects were transformed into a positive direction (i.e., treatment groups had better performance than control groups) and combined a synthetic effect size by means based on the suggestions of Borenstein et al. (2021).

$$d = \frac{\overline{X_1} - \overline{X_2}}{S_{pooled}}$$

The random-model method was employed to compute the overall effect size because these included SEL studies did not share a common true effect size but were sampled from a distribution of true effects. The true effect sizes varied from study to study since these studies had substantial differences in multiple aspects like methodological characteristics, implementation features, and program components, which were extracted and analyzed as moderators in the statistical process. The Q test (Borenstein et al., 2021) and I^2 (Higgins et al., 2003) were employed to estimate the heterogeneity. Multiple approaches, including the Classic fail-safe N test (Rosenthal, 1979), Orwin's fail-safe N test (Orwin, 1983), funnel plot, and subgroup analysis between published and unpublished studies, were used to examine publication bias.

$$Q = \sum_{i=1}^{k} W_i(Y_i - M)^2$$

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100\%$$

Regarding the moderator analysis, meta-regression was conducted to include all potential moderators in the regression model at the same time. Like the multiple

Table 2 Effect sizes of the 12 high-quality SEL programs

Domain	Number of studies	Overall ES (95% CI)	Social emotional skills ES	Affect and attitudes ES	Behavior ES	Academic performance ES
All programs	59	0.15* (0.11–0.19)	0.17*	0.09*	0.14*	0.13*
4Rs	2	0.10 (−0.05–0.24)	0.07	0.11	0.12*	0.02
Caring school community	2	0.15 (0.00–0.31)	0.18*	NA	0.20	0.09
Competent kids caring communities	1	0.22 (−0.15–0.58)	0.52*	0.36	NA	−0.07
I can problem solve	4	0.33* (0.11–0.55)	0.18*	0.23*	0.50*	NA
Lions quest	5	0.07* (0.03–0.11)	0.12*	0.10	0.07*	NA
MindUp	3	0.28* (0.04–0.52)	0.30*	0.18	0.44*	0.55*
Open circle	1	0.64* (0.31–0.98)	0.67*	NA	0.60*	NA
Positive action	5	0.40* (0.13–0.67)	0.30*	0.16	0.35*	0.34
PATHS	19	0.10* (0.06–0.15)	0.16*	0.06	0.06*	0.05*
RULER	2	0.07 (−0.07–0.21)	0.17	−0.05	0.06	0.37*
Second step	14	0.09* (0.04–0.15)	0.07	0.09*	0.09*	0.05*
Too good for violence	1	0.31* (0.18–0.45)	0.39*	0.07	0.29*	NA

$*p < 0.05$

regression approach in empirical studies, the meta-regression could help explore the relationship between study-level variables and effect sizes. Compared with the subgroup analysis that estimated only one moderator at one time, meta-regression could find the most effective program component after controlling for the impacts of methodological and implementation factors. In the meta-regression models, the overall effect and the four domains (i.e., social emotional skills, affect and attitudes, prosocial and anti-social behaviors, and academic performance) served as dependent variables separately; methodological characteristics, implementation features, and program components were used as categorical independent variables. In particular, Model 1 only included the five methodological predictors. Model 2 involved five methodological variables and one implementation variable. Potential program components coded as categorical variables were added to Model 3, Model 4, and Model 5 gradually. The interpretation of effect sizes was based on a set of empirical benchmarks for educational experimental research (small < 0.05, medium 0.05 to < 0.20, large ≥ 0.20; Kraft, 2020) All statistical analyses were performed using the software of Comprehensive Meta-Analysis version 3 (Borenstein et al., 2013).

Results

Overall Effects

The current review comprised 59 eligible studies and 463 effect sizes, involving 83,233 students (43,536 treatment, 39,697 control) from preschool to high school (see appendix 3 Table S3). Results showed that the overall effect size of these universal, high-quality, curriculum-based SEL programs was 0.15 ($CI = [0.11–0.19]$, $p < 0.01$). The effect sizes ranged from 0.13 to 0.16 after removing any one study, which lay in the 95% confidence interval, indicating that there was no outlier biasing the effects. The Q value was statistically significant, and I^2 value exceeded 75% (Higgins et al., 2003), indicating that these studies were highly heterogeneous ($Q = 326.49$, $df = 58$, $p < 0.01$; $I^2 = 82.24$). It is reasonable and feasible to employ moderators to explain the heterogeneity.

These SEL programs had the highest effect size on students' social emotional skills ($ES = 0.17$, $k = 44$). They also had significantly positive though small effects on affect and attitudes ($ES = 0.09$, $k = 24$), behaviors ($ES = 0.14$, $k = 43$) and academic performance ($ES = 0.13$, $k = 28$), indicating that SEL was effective. As to each SEL program, their overall effects varied significantly ($p < 0.01$). The effect sizes for Open Circle ($ES = 0.64$) and Positive Action ($ES = 0.40$) were relatively high, whereas the effects for Lions Quest, RULER and Second Step were minimal. Their effects on social emotional skills, affect and attitudes, prosocial and antisocial behaviors, and academic performance are shown in Table 2.

Results of Separate Program Components

A set of random-effects regression models was employed to explore the separate effects of each program component after controlling for methodological and implementation variables. The dependent variables were the overall effect

size and effect sizes for four outcome domains, while the five methodological and implementation variables (i.e., research design, sample size, grade, duration, and dosage) served as independent variables, and the five program components served as key independent variables separately. The results in Table 3 show the individual effects of each program component when added to the models with methodological and implementation factors. Results showed that cognitive elements and teacher social emotional skills significantly moderated the impact of SEL interventions, especially in the prosocial and antisocial behavior and academic performance domains. Pedagogical activities and climate support did not affect any SEL outcomes significantly. Family engagement was significantly related to the overall effects, behaviors, and academic performance. However, it was no longer significant when included in the regression models with cognitive elements and teacher social emotional skills. Only two program components (cognitive elements and teacher social emotional skills) were significant predictors of SEL effectiveness, whereas the other three (i.e., pedagogical activities, climate support, and family engagement) were not.

Meta-Regression for Overall Effects

To test the moderating effects of methodological characteristics, implementation features, and program components simultaneously, a set of random-effects regression models was conducted separately (Table 4). Methodological characteristics and implementation features were included in model 1 and model 2, acting as basic models without program components. Two program components, namely cognitive elements and teacher social emotional skills, were added in model 3, 4, and 5. Note that the other three insignificant program components (i.e., pedagogical activities, climate support, and family engagement) were not included in the models. In model 1, design and sample size were statistically significant, indicating that randomized controlled trials had lower effects and small sample studies had higher effects. In model 2, low dosage studies had significantly lower effects than studies with standard dosage. The coefficient of low dosage was the highest among all moderators, indicating that dosage was a predominant predictor of the effects. The studies that did not report dosage information also had significantly lower effects, which were similar to

Table 3 Results of program components in meta-regression (coefficient and standardized error)

Random effects	Overall ES	Social emotional skills	Affect and attitudes	Behavior	Academic performance
Cognitive elements	−0.102** (0.035)	−0.096† (0.052)	−0.059 (0.052)	−0.110** (0.042)	−0.241** (0.078)
Teacher social emotional skills	0.137*** (0.041)	0.044 (0.058)	0.068 (0.050)	0.168*** (0.047)	0.199** (0.062)
Activities	−0.016 (0.042)	−0.009 (0.059)	−0.020 (0.054)	−0.020 (0.047)	0.026 (0.086)
Climate	0.014 (0.036)	0.037 (0.044)	0.077† (0.042)	0.051 (0.045)	0.071 (0.080)
Family engagement	0.087* (0.035)	0.047 (0.053)	0.067 (0.045)	0.086* (0.041)	0.162* (0.069)

$^\dagger p < 0.1$, $* < 0.05$, $**p < 0.01$, $***p < 0.001$

Table 4 Results of meta-regression for overall effects (coefficient and standardized error)

Random effects	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	0.190*** (0.032)	0.234*** (0.032)	0.297*** (0.037)	0.216*** (0.029)	0.262*** (0.037)
Design (RCT)	−0.129** (0.041)	−0.120** (0.040)	−0.108** (0.037)	−0.137*** (0.035)	−0.124*** (0.036)
Sample size (S)	0.152* (0.070)	0.125† (0.068)	0.128† (0.066)	0.082 (0.066)	0.095 (0.066)
Grade (Pre)	0.066 (0.064)	0.088 (0.059)	0.117* (0.056)	0.111* (0.054)	0.125* (0.054)
Grade (Secondary)	−0.021 (0.059)	−0.030 (0.052)	−0.030 (0.048)	−0.003 (0.046)	−0.011 (0.045)
Duration (more than 1 year)	−0.007 (0.042)	0.053 (0.041)	0.070† (0.038)	0.037 (0.036)	0.053 (0.037)
Dosage (low)		−0.141*** (0.040)	−0.154*** (0.037)	−0.123*** (0.035)	−0.136*** (0.035)
Dosage (NA)		−0.125* (0.053)	−0.139** (0.050)	−0.123** (0.047)	−0.132** (0.047)
Cognitive elements			−0.102** (0.035)		−0.068† (0.036)
Teacher social emotional skills				0.137*** (0.041)	0.104* (0.044)
Q	28.34	47.72	62.65	69.98	75.43
df	5	7	8	8	9
R^2 analog	0.20	0.42	0.54	0.61	0.63

$^\dagger p < 0.1$, $* < 0.05$, $**p < 0.01$, $***p < 0.001$

Table 5 Results of meta-regression for multiple outcomes (coefficient and standardized error)

Random effects	Overall ES	Social emotional skills	Affect and attitudes	Behavior	Academic performance
Intercept	0.262*** (0.037)	0.335*** (0.064)	0.122* (0.058)	0.256*** (0.042)	0.311** (0.100)
Design (RCT)	−0.124*** (0.036)	−0.076 (0.059)	−0.070 (0.062)	−0.113** (0.039)	−0.102 (0.093)
Sample size (S)	0.095 (0.066)	0.188* (0.095)	0.137 (0.118)	0.097 (0.081)	0.078 (0.123)
Grade (Pre)	0.125* (0.054)	0.100 (0.075)	0.172† (0.096)	0.116 (0.073)	0.040 (0.111)
Grade (Secondary)	−0.011 (0.045)	–	–	−0.009 (0.047)	0.060 (0.146)
Duration (more than 1 year)	0.053 (0.037)	0.026 (0.061)	0.113 (0.075)	0.044 (0.042)	−0.011 (0.086)
Dosage (low)	−0.136*** (0.035)	−0.191*** (0.057)	−0.109 (0.068)	−0.129*** (0.039)	−0.105 (0.067)
Dosage (NA)	−0.132** (0.047)	−0.137† (0.072)	−0.062 (0.062)	−0.135* (0.057)	−0.122 (0.117)
Cognitive elements	−0.068† (0.036)	−0.094† (0.056)	−0.040 (0.055)	−0.068† (0.041)	−0.130 (0.100)
Teacher social emotional skills	0.104* (0.044)	0.007 (0.061)	0.053 (0.054)	0.130* (0.053)	0.136† (0.077)
k	59	44	24	43	28
Q	75.43	40.28	12.49	54.31	33.57
df	9	8	8	9	9
R^2 analog	0.63	0.42	0.23	0.60	0.47

†$p < 0.1$, *< 0.05, **< 0.01, ***$p < 0.001$

the low dosage studies. Cognitive elements and teacher social emotional skills were added in model 3 and model 4 respectively, and their coefficients were both significant. In model 5, these two variables were added to the equation simultaneously. The coefficient of teacher social emotional skills was still significant, whereas the coefficient of cognitive elements was marginally significant, indicating that both were effective predictors of the program effects. As to the R^2 analog, it increased as the number of variables increased. Taking R^2 analog and model simplicity into consideration, model 5 was chosen as the final model. In this model, the coefficients of design, grade, dosage, cognitive elements, and teacher social emotional skills were statistically significant, which was roughly consistent among all models. The R^2 analog of model 5 reached 63%, showing that the predictors could explain a large proportion of the between-studies variance. The coefficient of low dosage was largest ($ES = −0.14$), indicating its highest relation with SEL outcomes. As to the program components, teacher social emotional skills could significantly improve the effectiveness of SEL programs ($ES = 0.10$, $p < 0.05$), whereas cognitive elements could reduce their effectiveness ($ES = −0.07$, $p < 0.1$).

Meta-Regression for Multiple Outcomes

Since the program outcomes were classified into four domains, a set of regression models of methodological characteristics, implementation features, and program components was conducted to explain the variance on different outcomes (Table 5). These models were parallel to

the above model 5 of overall effects, and their results were relatively consistent with the exception of effect and attitudes. The distinct results may be due to the small number of included studies involving affect and attitudes outcomes. The included studies with affect and attitudes outcomes were not highly heterogeneous ($Q = 64.29$, $df = 23$, $p < 0.01$; $I^2 = 64.23\% < 0.75$), which may also contribute to the insignificant moderation effects.

When comparing these models, it could be found that after controlling for the methodological and implementation variables, cognitive elements and teacher social emotional skills could significantly moderate the program effects in terms of social emotional skills, behaviors and academic performance. The results suggested that focusing on cognitive elements was negatively associated with SEL outcomes regarding social-emotional skills and prosocial behaviors ($ES = −0.09$, $p < 0.1$; $ES = −0.07$, $p < 0.1$). Cognitive elements had no significant impact on other SEL outcomes, including affect and attitudes and academic performance. A focus on teacher social emotional skills was significantly associated with increased effectiveness in terms of behaviors and academic performance ($ES = 0.13$, $p < 0.05$; $ES = 0.14$, $p < 0.1$), but not with social emotional skills and affect and attitudes domains. Low dosage was an important factor related to the program effects, especially social emotional skills and behaviors. In short, both cognitive elements and teacher social emotional skills were significant predictors of SEL programs after controlling for methodological and implementation factors. The moderation effects of cognitive elements were mainly negative whereas those of teacher social emotional skills were positive.
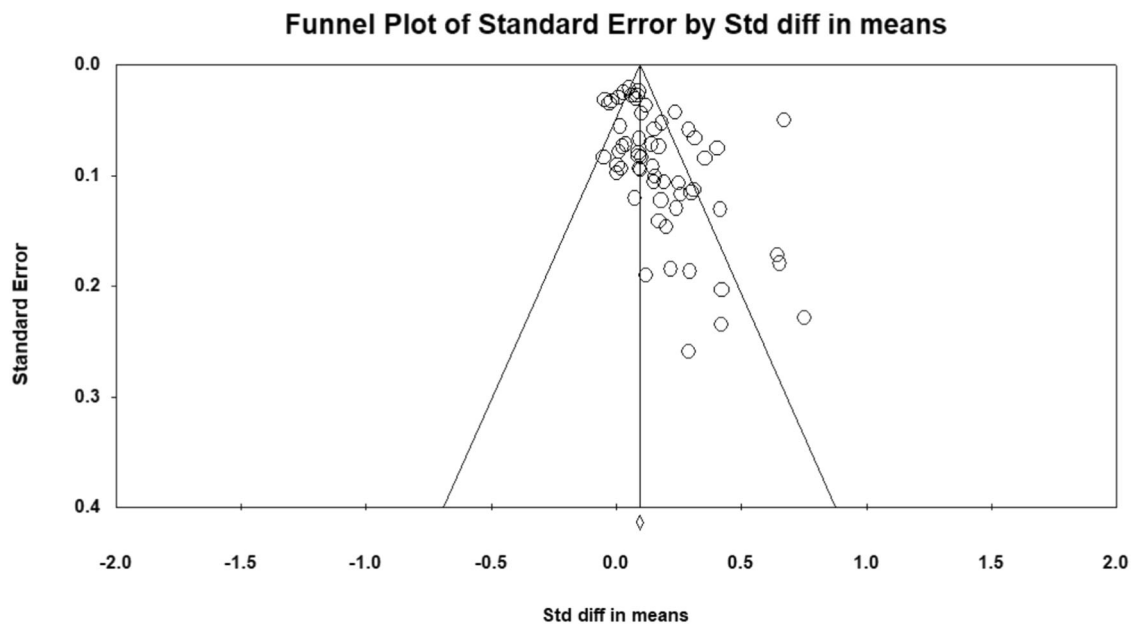
Fig. 2 Funnel plot

Publication Bias

Multiple approaches were employed to examine publication bias. The result of Classic fail-safe N test showed that 3634 missing studies were required to make the effects become zero. Orwin's fail-safe N test showed that 489 missing studies were required if the trivial value was set to the 0.01 level. Because these requirements were difficult to achieve, it could be concluded that there was no publication bias. Funnel plot was employed to exhibit the distribution of effect sizes intuitively, indicating that the effect sizes were not completely symmetrical (Fig. 2). However, adding publication status to the final regression models was tried, and its effects were not significant, indicating that there was no publication bias.

Discussion

A set of reviews has examined the impacts of SEL programs on youth and adolescents, but none of them has employed a meta-analytic approach to identify the core program components and related methodological and implementation moderators that make programs effective. Based on the comparison of previous high-quality SEL programs, this meta-analytic review identified five key program components that might affect SEL effectiveness, namely cognitive elements, pedagogical activities, teacher social emotional skills, climate support, and family engagement. The current study chose 12 universal high-quality curriculum-based SEL programs and tested their effects on social emotional skills, affect and attitudes, prosocial and antisocial

behaviors, and academic performance. It further examined the effects of the five program components and other methodological and implementation moderators (i.e., research design, sample size, grade, duration, dosage) on SEL effectiveness. Meta-regression results indicated that training teacher social emotional skills and reducing cognitive elements in curricula could produce better SEL outcomes, whereas pedagogical activities, climate support, and family engagement could not.

Collectively, 59 eligible SEL studies involving 83,233 participants were included, indicating a positive and small effect ($ES = 0.15$). These universal, high-quality, curriculum-based SEL programs significantly improved students' social emotional skills ($ES = 0.17$, $k = 44$), reinforced affect and attitudes ($ES = 0.09$, $k = 24$), promoted academic performance ($ES = 0.13$, $k = 28$), and increased prosocial behaviors and reduced antisocial behaviors ($ES = 0.14$, $k = 43$). Compared with some previous reviews, the overall effects and sub-domain effects in the current review were relatively smaller. The effects of SEL on multiple outcomes ranged from 0.09 to 0.70 in previous meta-analyses (Durlak et al., 2011; Sklad et al., 2012; Wigelsworth et al., 2016). One determinant reason for the distinct effects might be the stringent inclusion criteria employed in the current review, which excluded a lot of low-quality studies or small-scale studies with high effect sizes. In previous meta-analyses, small studies tended to generate higher effects (Cheung & Slavin, 2016; Slavin & Smith, 2009). 47 out of the 59 included studies in the current review had a relatively large number of participants, leading to smaller effect sizes. Based on the suggestions for interpreting effects of education

interventions by Kraft (2020), the effect sizes in the current review were medium and meaningful.

The current review provided an overview of the effects of twelve high-quality, curriculum-based SEL programs on Pre-K-12 students in terms of social emotional skills, affect and attitudes, prosocial and antisocial behaviors, and academic performance. Open Circle had the largest effect size of 0.64 among all the programs, but it only had one quasi-experimental study with 147 students, reducing the credibility of its effects. Positive Action had an overall effect size of 0.40, and the corresponding five studies involved both quasi-experimental studies and randomized controlled studies, large studies and small studies, preschool studies and elementary and secondary studies. The numbers of included studies of PATHS and Second Step were relatively large, but their effects were relatively small. As to the effects on each domain, Open Circle had the largest effect on social emotional skills and behaviors, Competent Kids Caring Communities had the largest effect on affect and attitudes, and MindUP had the largest effect on academic performance. A potential explanation for the high effectiveness of Open Circle could be its low proportion of cognitive elements in the curriculum and strong emphasis on promoting teachers' social emotional skills in teacher training. For a more comprehensive understanding of the 12 high-quality SEL programs, please refer to Table S2 in the appendix for a brief description of each program. Although these programs are well-known high-quality SEL programs, their effect sizes varied significantly in terms of overall effects and four outcome domains.

Regarding the moderators, methodological characteristics, implementation features and program components were highly related to the effects of these programs, which explained 63% of between-studies variance. First, randomized controlled trials had a significantly lower effect than quasi-experimental studies, which was consistent with the previous reviews. For instance, Cheung and Slavin (2016) explored the influences of methodological features on the effect sizes in educational programs, and found that the effects of randomized controlled trials were significantly lower than those of quasi-experimental studies. Second, preschool studies had a higher effect than elementary studies, which was similar to the effects of Second Step on Pre-Kindergarten students in terms of antisocial behaviors (Moy et al., 2018). This result was partially consistent with one previous SEL review, which found that mean age negatively moderated the SEL effects in terms of social emotional skills (Durlak et al., 2011). The moderation effects of sample size and duration were not significant. Finally, dosage was examined to be a predominant implementation feature. Studies with low dosages had significantly smaller effects than studies with standard dosages, indicating that sufficient dosage was a determinant of SEL effectiveness

even though they were all high-quality programs. The results further verified and extended the crucial impacts of implementation features, especially the dosage proposed by previous studies (e.g., Shi et al., 2022; Yang et al., 2019).

As to the core program components, cognitive elements and teacher social emotional skills were both significant predictors but in opposite directions. After controlling for methodological features and implementation factors, highlighting teacher social emotional skills in teacher training significantly improved the overall effects of SEL programs ($ES = 0.10$, $p < 0.05$), whereas increasing the proportion of cognitive elements in curricula reduced SEL effects ($ES = -0.07$, $p < 0.1$). Regarding different outcome domains, teacher social emotional skills improved SEL outcomes in behaviors and academic performance but not in social emotional skills and affect and attitudes. The significant moderating effect of teacher social emotional skills roughly coincided with the prosocial classroom model, which pointed out that teacher social emotional competence was a crucial antecedent to students' social, emotional, and academic outcomes (Jennings & Greenberg, 2009). The partially insignificant results might be influenced by the limited number of included studies in these domains or other potential moderators, which warrant further investigation. The inclusion of cognitive elements in SEL curricula was found to have a negative impact on social emotional skills and behaviors, but not on academic performance and affect and attitudes domains. This is consistent with previous research that has found a marginally significantly negative influence of cognitive elements on peer relationships (Cipriano et al., 2023). Cognitive elements might crowd out the scarce and precious time of SEL curricula, leading to reduced SEL effects. Integrating cognitive elements into SEL did not significantly affect youth academic performance, consistent with previous meta-analytic results that academic integration did not moderate the impact of SEL on school functioning such as academic achievement (Cipriano et al., 2023). One possible reason for the insignificant effect of cognitive elements on academic performance was that, compared with the daily cognitive instructions in schools, the role of infiltrating cognitive elements in SEL was negligible. Insignificant results concerning the affect and attitudes domain were perhaps due to the incongruent grouping of outcomes. The findings also suggest that pedagogical activities, climate support, and family engagement did not significantly affect SEL effectiveness. Although some studies supported that they might be potential moderators, they seemed to be complementary modules compared to the two significant program components (i.e., cognitive elements in curricula, teacher social emotional skills in teacher training) in comprehensive SEL programs. The presence of other methodological factors and implementation features might interfere with the moderating effects of pedagogical activities, climate support, and family

engagement to some extent. Therefore, to enhance effectiveness, SEL programs should pay more attention to training teacher social emotional skills and may consider reducing cognitive elements in course content.

There were three essential contributions of the current review. First, we found that large-scale studies of SEL programs tended to generate smaller effect sizes in terms of social emotional skills and other outcomes. It might suggest that SEL programs were difficult to scale, even if they were well-designed and provided multiple auxiliary components. The included twelve SEL programs, for example, were all high-quality and widely-used by practitioners and researchers, yet they had small effects when applied to larger populations. This finding is consistent with sibling programs or reform efforts, which are also hard to scale. For instance, Cheung and Slavin (2016) found that the larger the scale of reading, mathematics and science programs, the smaller their effects. How to maintain the effects of small experimental studies while scaling them up is a common and crucial issue for educational practice.

Relatedly, in order to reinforce the wide-scale adoption of comprehensive SEL programs as well as other sibling programs, it is important to understand which components should be scaled, or which components are effective to the outcomes. The current review stands in this direction, indicating that training teacher social emotional skills and reducing cognitive elements were significant predictors of SEL outcomes, whereas pedagogical activities, climate support, and family engagement were not. The findings about teacher social emotional skills training were consistent with the previous descriptions, which mentioned that preparing teachers with necessary social emotional skills is crucial to the effectiveness and scaling up of SEL programs (Elias et al., 2003). The current review is the first to provide evidence for this statement. We further examined the teacher social emotional training practices in the corresponding programs, namely 4Rs, Positive Action, Ruler and MindUp, and found that they all emphasized ongoing coaching. For instance, the ongoing coaching of 4Rs program provided group meetings, workshops, lesson modeling to help teachers reinforce their teaching and promote their professional development (Jones et al., 2010). Ongoing coaching might be a promising strategy in teacher social emotional training to maintain the effects of large-scale SEL programs. Further studies should be conducted to examine the effects of ongoing coaching as well as other teacher professional development strategies.

The last contribution was the innovation in research method. To the best of our knowledge, there was no previous review that examined the effective components by conducting a meta-analysis of some specific programs. Most comprehensive meta-analyses often include all accessible SEL programs (Durlak et al., 2011), whereas SEL guides aimed to examine program components always summarize qualitative information systematically (Jones et al., 2017). One recent review of student-teacher relationships noticed this issue and conducted meta-analysis and common components approach separately, but it did not code the components quantitatively to be analyzed in meta-analysis (Kincade et al., 2020). The current review provided a new quantitative approach to examine effective components in programs, in which components were summarized as moderators and then analyzed in meta-analysis.

The current review still had limitations. First, some included programs only had one or two studies, including but not limited to the Competent Kids Caring Communities program and the Open Circle program, which reduced the reliability of their effect sizes. As the purpose of the current review was to find the components moderating the program outcomes, the small numbers of included studies involving these programs had almost no influence on identifying effective components and examining moderation effects. Second, each outcome domain seemed to be inclusive, which might bias the results. For instance, the concepts of self-concept, empathy and social skill may be somewhat different, but we classified them into the same domain of social emotional skills. This loose categorization may affect the results to some extent and the findings should be interpreted with caution. It is difficult to classify the outcomes concisely owing to the variety of SEL outcomes. Our categorization was roughly consistent with previous SEL reviews (e.g., Durlak et al., 2011), suggesting that the current categorization was reasonable and acceptable. Third, there are some potential issues in statistical analyses (i.e., considering $p < 0.1$ as marginally significant, coding all moderators as dichotomic variables), which may make the results fragile. The current findings should be interpreted with caution. There may be other moderators related to SEL effectiveness that could bias the results, which should be further explored in future studies. Since most cost-benefit analyses focus on programs rather than components, if the costs of program components could be taken into consideration as well as their effects, it could provide more evidence-based effective and economical suggestions for SEL practice.

Conclusion

Previous reviews have yielded valuable insights into the effectiveness of SEL programs. However, a limitation has been the absence of a meta-analytic approach with stringent criteria and diverse moderators to identify the key components that make these programs work. To facilitate the generalization of SEL programs, it is imperative to figure out which program components warrant scaling and which ones influence the outcomes. Based on two fundamental reports, namely CASEL (2013) and Jones et al. (2017), the

present study selected 12 universal, school-based, and high-quality SEL programs. The study systematically synthesized the empirical evidence of their effectiveness, yielding a significant positive and small effect. Furthermore, the study identified the key program components capable of significantly moderating SEL program outcomes while controlling for the methodological and implementation features. The findings indicated that cognitive elements and teacher social emotional skills were supported as effective components. These results imply that optimizing SEL programs might benefit from enhancing teacher training in social emotional skills and reducing the cognitive content of curricula to achieve effectiveness. The findings revealed a significant negative impact of low dosage on SEL effects, thereby underscoring the importance of ensuring a sufficient dosage of SEL interventions, regardless of program quality. In summary, this review presents a comprehensive overview of the impact of SEL programs on multiple outcomes, including social emotional skills, affect and attitudes, pro-social and antisocial behaviors, and academic performance. It highlights the importance of teacher social emotional skills and the imperative of ensuring sufficient dosage when implementing SEL programs. The study sheds light on how to maintain the effectiveness of SEL interventions when scaling them up and provides evidence-based recommendations for the practical implementation of SEL within educational contexts.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Compliance with ethical standards

Conflict of Interest The authors declare no competing interests.

Ethical Approval This research does not contain any studies with human participants conducted by any of the authors.

References

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

Aras, C. Y., & Aslan, D. (2018). The effects of" I can problem solve program" on children's perspective taking abilities. *International Journal of Evaluation and Research in Education*, 7(2), 109–117.

Averdijk, M., Zirk-Sadowski, J., Ribeaud, D., & Eisner, M. (2016). Long-term effects of two childhood psychosocial interventions on adolescent delinquency, substance use, and antisocial behavior: A cluster randomized controlled trial. *Journal of Experimental Criminology*, 12(1), 21–47. https://doi.org/10.1007/s11292-015-9249-4.

Berkel, C., Mauricio, A. M., Schoenfelder, E. & & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science*, 12, 23–33. https://doi.org/10.1007/s11121-010-0186-1.

Blewitt, C., Fuller-Tyszkiewicz, M., Nolan, A., Bergmeier, H., Vicary, D., Huang, T., McCabe, P., McKay, T., & Skouteris, H. (2018). Social and emotional learning associated with universal curriculum-based interventions in early childhood education and care centers a systematic review and meta-analysis. *Jama Network Open*, 1(8), e185727 https://doi.org/10.1001/jamanetworkopen.2018.5727.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2013). *Comprehensive meta-analysis* (Version 3) [Computer software]. Biostat. https://www.meta-analysis.com/

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis* (2nd ed.). John Wiley & Sons.

Brackett, M. A., Rivers, S. E., Reyes, M. R., & Salovey, P. (2012). Enhancing academic performance and social and emotional competence with the RULER feeling words curriculum. *Learning and Individual Differences*, 22(2), 218–224. https://doi.org/10.1016/j.lindif.2010.10.002.

Bradshaw, C. P., Shukla, K. D., Pas, E. T., Berg, J. K., & Ialongo, N. S. (2020). Using complier average causal effect estimation to examine student outcomes of the pax good behavior game when integrated with the PATHS curriculum. *Administration and Policy in Mental Health and Mental Health Services Research*, 47(6), 972–986. https://doi.org/10.1007/s10488-020-01034-1.

Cipriano, C., Strambler, M. J., Naples, L. H., Ha, C., Kirk, M., Wood, M., & Durlak, J. (2023). The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions. *Child Development*, 94(5), 1181–1204. https://doi.org/10.1111/cdev.13968.

Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292.

Cohen, J. (1987). *Statistical power analysis for the behavioral sciences*. Hillside, NJ: Lawrence Erlbaum Associates.

CASEL. (2013). *2013 CASEL guide: Effective Social and Emotional Learning Programs-Preschool and Elementary School Edition*. https://casel.org/wp-content/uploads/2016/01/2013-casel-guide-1.pdf

CASEL. (2015). *2015 CASEL guide: Effective Social and Emotional Learning Programs-Middle and High School Edition*. http://secondaryguide.casel.org/casel-secondary-guide.pdf

Corcoran, R. P., Cheung, A. C. K., Kim, E., & Xie, C. (2018). Effective universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research Review*, 25, 56–72. https://doi.org/10.1016/j.edurev.2017.12.001.

Duncan, R., Washburn, I. J., Lewis, K. M., Bavarian, N., DuBois, D. L., Acock, A. C., Vuchinich, S., & Flay, B. R. (2017). Can

Universal SEL programs benefit universally? Effects of the positive action program on multiple trajectories of social-emotional and misconduct behaviors. *Prevention Science*, 18(2), 214–224. https://doi.org/10.1007/s11121-016-0745-1.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327–350. https://doi.org/10.1007/s10464-008-9165-0.

Elias, M. J., Zins, J. E., Graczyk, P. A., & Weissberg, R. P. (2003). Implementation, sustainability, and scaling up of social-emotional and academic innovations in public schools. *School Psychology Review*, 32(3), 303–319. https://doi.org/10.1080/02796015.2003.12086200.

Farrell, A. D., Sullivan, T. N., Goncy, E. A., & Le, A. T. H. (2016). Assessment of adolescents' victimization, aggression, and problem behaviors: Evaluation of the problem behavior frequency scale. *Psychological Assessment*, 28(6), 702–714. https://doi.org/10.1037/pas0000225.

Fishbein, D. H., Domitrovich, C., Williams, J., Gitukui, S., Guthrie, C., Shapiro, D., & Greenberg, M. (2016). Short-term intervention effects of the paths curriculum in young low-income children: Capitalizing on plasticity. *Journal of Primary Prevention*, 37(6), 493–511. https://doi.org/10.1007/s10935-016-0452-5.

Flay, B. R., & Allred, C. G. (2003). Long-term effects of the Positive Action® program. *American Journal of Health Behavior*, 27(1), S6–S21.

Frey, K. S., Hirschstein, M. K., & Guzzo, B. A. (2000). Second step: Preventing aggression by promoting social competence. *Journal of Emotional and Behavioral Disorders*, 8(2), 102–112. https://doi.org/10.1177/106342660000800206.

Greenberg, M. T., & Kusché, C. A. (2006). Building social and emotional competence: The PATHS curriculum. In S. R. Jimerson & M. Furlong (Eds.), *Handbook of school violence and school safety: From research to practice* (pp. 395–412). Lawrence Erlbaum Associates Publishers.

Gresham, F. M., & Elliott, S. N. (1990). *The Social Skills Rating System*. Circle Pines, MN: American Guidance Service.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411–482.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128. https://doi.org/10.3102/10769986006002107.

Hennessey, B. A. (2007). Promoting social competence in school-aged children: The effects of the open circle program. *Journal of School Psychology*, 45(3), 349–360. https://doi.org/10.1016/j.jsp.2006.11.007.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557.

Huebner, E. S. (1991). Initial development of the student's life satisfaction scale. *School Psychology International*, 12(3), 231–240. https://doi.org/10.1177/0143034391123010.

Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research*, 79(1), 491–525. https://doi.org/10.3102/0034654308325693.

Jones, S., Brush, K., Bailey, R., Brion-Meisels, G., McIntyre, J., Kahn, J., Nelson, B., & Stickle, L. (2017). *Navigating social and emotional learning from the inside out; Looking inside and across 25 leading SEL programs: A practical resource for schools and OST providers (elementary school focus)*. Harvard Graduate School of Education. https://www.wallacefoundation.org/knowledge-center/Documents/Navigating-Social-and-Emotional-Learning-from-the-Inside-Out.pdf

Jones, S. M., & Bouffard, S. M. (2012). Social and emotional learning in schools: From programs to strategies and commentaries. *Social Policy Report*, 26(4), 1–33.

Jones, S. M., Brown, J. L., & Aber, J. L. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development*, 82(2), 533–554. https://doi.org/10.1111/j.1467-8624.2010.01560.x.

Jones, S. M., Brown, J. L., Hoglund, W. L. G., & Aber, J. L. (2010). A school-randomized clinical trial of an integrated social-emotional learning and literacy intervention: Impacts after 1 school year. *Journal of Consulting and Clinical Psychology*, 78(6), 829–842. https://doi.org/10.1037/a0021383.

Kelty, N. E., & Wakabayashi, T. (2020). Family engagement in schools: Parent, educator, and community perspectives. *SAGE Open*, 10(4). https://doi.org/10.1177/2158244020973024

Kincade, L., Cook, C., & Goerdt, A. (2020). Meta-analysis and common practice elements of universal approaches to improving student-teacher relationships. *Review of Educational Research*, 90(5), 710–748. https://doi.org/10.3102/0034654320946836.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253.

Laurent, J., Catanzaro, S. J., Joiner, Jr, T. E., Rudolph, K. D., Potter, K. I., Lambert, S., Osborne, L., & Gathright, T. (1999). A measure of positive and negative affect for children: Scale development and preliminary validation. *Psychological Assessment*, 11(3), 326–338. https://doi.org/10.1037/1040-3590.11.3.326.

Lawson, G. M., McKenzie, M. E., Becker, K. D., Selby, L., & Hoover, S. A. (2019). The core components of evidence-based social emotional learning programs. *Prevention Science*, 20(4), 457–467.

LeBuffe, P. A., Shapiro, V. B., & Naglieri, J. A. (2009). *The Devereux student strengths assessment (DESSA): Assessment, technical manual, and user's guide*. Charlotte, NC: Apperson, Inc.

Low, S., Smolkowski, K., Cook, C., & Desfosses, D. (2019). Two-year impact of a universal social-emotional learning curriculum: Group differences from developmentally sensitive trends over time. *Developmental Psychology*, 55(2), 415–433. https://doi.org/10.1037/dev0000621.

Miyamoto, K., Huerta, M. C., & Kubacka, K. (2015). Fostering social and emotional skills for well-being and social progress. *European Journal of Education*, 50(2), 147–159.

Moy, G., Polanin, J. R., McPherson, C., & Thuy-Vy, P. (2018). International adoption of the Second Step program: Moderating variables in treatment effects. *School Psychology International*, 39(4), 333–359. https://doi.org/10.1177/0143034318783339.

Moy, G. E., & Hazen, A. (2018). A systematic review of the Second Step program. *Journal of School Psychology*, 71, 18–41. https://doi.org/10.1016/j.jsp.2018.10.006.

Murano, D., Sawyer, J. E., & Lipnevich, A. A. (2020). A meta-analytic review of preschool social and emotional learning interventions. *Review of Educational Research*, 90(2), 227–263. https://doi.org/10.3102/0034654320914743.

Neitzel, A. J., Lake, C., Pellegrini, M., & Slavin, R. E. (2022). A synthesis of quantitative research on programs for struggling readers in elementary schools. *Reading Research Quarterly*, 57(1), 149–179. https://doi.org/10.1002/rrq.379.

Orpinas, P., & Frankowski, R. (2001). The aggression scale: A self-report measure of aggressive behavior for young adolescents. *The

Journal of Early Adolescence, 21(1), 50–67. https://doi.org/10.1177/0272431601021001003.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, 8(2), 157–159.

Rosenberg, M. (1989). *Society and the adolescent self-image. (Rev. ed.)*. Middletown, CT: Wesleyan University Press.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638.

Sabey, C. V., Charlton, C. T., Pyle, D., Lignugaris-Kraft, B., & Ross, S. W. (2017). A review of classwide or universal social, emotional, behavioral programs for students in kindergarten. *Review of Educational Research*, 87(3), 512–543.

Sandilos, L., Goble, P., & Schwartz, S. (2020). Burnout and teacher-child interactions: The moderating influence of SEL interventions in head start classrooms. *Early Education and Development*, 31(7), 1169–1185. https://doi.org/10.1080/10409289.2020.1788331.

Schonfeld, D. J., Adams, R. E., Fredstrom, B. K., Weissberg, R. P., Gilman, R., Voyce, C., Tomlin, R., & Speese-Linehan, D. (2015). Cluster-randomized trial demonstrating impact on academic achievement of elementary social-emotional learning. *School Psychology Quarterly*, 30(3), 406–420. https://doi.org/10.1037/spq0000099.

Shi, J., Cheung, A. C., & Ni, A. (2022). The effectiveness of promoting alternative thinking strategies program: A meta-analysis. *Frontiers in Psychology*, 13, 1030572. https://doi.org/10.3389/fpsyg.2022.1030572.

Sklad, M., Diekstra, R., De Ritter, M., Ben, J., & Gravesteijn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and adjustment? *Psychology in the Schools*, 49(9), 892–909. https://doi.org/10.1002/pits.21641

Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis*, 31(4), 500–506.

Stanley, L. (2019). Is the preschool PATHS curriculum effective? A review. *Journal of Evidence-Based Social Work*, 16(2), 130–143.

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, 88(4), 1156–1171. https://doi.org/10.1111/cdev.12864.

What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook, version 4.1*. National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf

Wigelsworth, M., Lendrum, A., Oldfield, J., Scott, A., ten Bokkel, I., Tate, K., & Emery, C. (2016). The impact of trial stage, developer involvement and international transferability on universal social and emotional learning programme outcomes: A meta-analysis. *Cambridge Journal of Education*, 46(3), 347–376. https://doi.org/10.1080/0305764x.2016.1195791.

Wigelsworth, M., Mason, C., Verity, L., Qualter, P., & Humphrey, N. (2022). Making a case for core components: New frontiers in SEL theory, research, and practice. School Psychology Review, 1–14. https://doi.org/10.1080/2372966X.2021.2004863

Wilson, B. C. (2016). *The Effectiveness of Promoting Alternative Thinking Strategies (PATHS) When Used One Time per Week in Therapeutic Day Treatment*. (Publication Number 10096120) [Doctoral dissertation, Walden University]. ProQuest Dissertations & Theses A&I.

Wu, Z., Brown, L., Kim, H. Y., Yoshikawa, H., & Aber, J. L. (2023). Measuring the dosage of brief and skill-targeted social-emotional learning (SEL) activities in humanitarian settings. *Frontiers in Psychology*, 13, 973184. https://doi.org/10.3389/fpsyg.2022.973184.

Yang, W., Datu, J. A. D., Lin, X., Lau, M. M., & Li, H. (2019). Can early childhood curriculum enhance social-emotional competence in low-income children? A meta-analysis of the educational effects. *Early Education and Development*, 30(1), 36–59. https://doi.org/10.1080/10409289.2018.1539557.

Jieping Shi is an assistant professor in the Faculty of Education at Beijing Normal University. Her research interests are social emotional learning, quantitative methods, meta-analysis, and evidence-based reform.

Alan C.K. Cheung is the chairperson and professor in the Department of Educational Administration and Policy at The Chinese University of Hong Kong. His research interests include meta-analysis, evidence-based reform, and large-scale experimental studies.