ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/42346373

# A Meta-Analysis of After-School Programs That Seek to Promote Personal and Social Skills in Children and Adolescents

**Article** *in* American Journal of Community Psychology · March 2010

DOI: 10.1007/s10464-010-9300-6 · Source: PubMed

CITATIONS
758

READS
6,184

**3 authors**, including:



Joseph A Durlak
Loyola University Chicago
**144** PUBLICATIONS **19,880** CITATIONS

SEE PROFILE



Roger P. Weissberg
University of Illinois at Chicago
**140** PUBLICATIONS **16,531** CITATIONS

SEE PROFILE

**Some of the authors of this publication are also working on these related projects:**



Positive Youth Development View project



social and emotional learning View project

All content following this page was uploaded by Joseph A Durlak on 15 May 2014.

The user has requested enhancement of the downloaded file.

Am J Community Psychol (2010) 45:294–309
DOI 10.1007/s10464-010-9300-6

ORIGINAL PAPER

# A Meta-Analysis of After-School Programs That Seek to Promote Personal and Social Skills in Children and Adolescents

**Joseph A. Durlak · Roger P. Weissberg · Molly Pachan**

Published online: 19 March 2010
© Society for Community Research and Action 2010

**Abstract** A meta-analysis of after-school programs that seek to enhance the personal and social skills of children and adolescents indicated that, compared to controls, participants demonstrated significant increases in their self-perceptions and bonding to school, positive social behaviors, school grades and levels of academic achievement, and significant reductions in problem behaviors. The presence of four recommended practices associated with previously effective skill training (SAFE: sequenced, active, focused, and explicit) moderated several program outcomes. One important implication of current findings is that ASPs should contain components to foster the personal and social skills of youth because youth can benefit in multiple ways if these components are offered. The second implication is that further research is warranted on identifying program characteristics that can help us understand why some programs are more successful than others.

**Keywords** After-school · Meta-analysis · Social competence · Social skills · Youth development

## Introduction

What is known about the impact of after-school programs (ASPs)? Considerable attention has focused on the academic benefits of ASPs. The results of two large-scale evaluations of twenty-first Century Community Learning Centers (21 CCLCs) that is, centers that received federal funding through No Child Left Behind legislation have generated controversy. Neither the evaluation of centers serving elementary (James-Burdumy et al. 2005) or middle school students (Dynarski et al. 2004) found any significant gains in achievement test scores, although there were some gains in secondary outcomes such as parental involvement in school and student commitment to homework. These findings led some to suggest drastic reductions in the levels of federal financial support for ASPs, which had reached one billion dollars a year by 2002 (Mahoney and Zigler 2006).

However, researchers have discussed several methodological issues that limit the interpretation of the results of the national evaluations of 21 CCLCs (Kane 2004; Mahoney and Zigler 2006). Depending on the age group in question, these include the lack of initial group equivalence, high attrition among respondents, low levels of student attendance, and the possible nonrepresentativeness of evaluated programs. There is also the problem of treating centers as though they provided a uniform approach to academic assistance when they clearly did not. While some 21 CCLCs provided students with intensive small group instruction or individual tutoring, others merely asked students to work independently on homework.

Instead of focusing on the results of only two evaluations out of many, a well-done meta-analysis that evaluates a broad sample of relevant studies carefully can assess the magnitude of change on different outcomes and identify some of the important characteristics of programs associated with more positive results. For example, the meta-analysis of 35 outcome studies by Lauer et al. (2006) led to the conclusion that ASPs "…can have positive effects on the achievement of academically at-risk students" (p. 303).

J. A. Durlak (✉) · M. Pachan
Department of Psychology, Loyola University Chicago, 6525 N. Sheridan Road, Chicago, Il 60626, USA
e-mail: jdurlak@luc.edu

R. P. Weissberg
Department of Psychology, University of Illinois at Chicago & Collaborative for Academic, Social, and Emotional Learning (CASEL), Chicago, Il, USA

Significant gains in reading or math achievement, or in both areas, were observed for elementary, middle, and high school students, and the latter group showed the most improvement in both areas. Although the results of a meta-analysis are never definitively conclusive, Lauer et al.'s (2006) results begin to clarify which program participants might be more likely to derive academic benefits from ASPs.

## What About Personal and Social Benefits?

The recent focus on the academic benefits of ASPs tends to overlook the fact that many ASPs were initially created based on the idea that young people's participation in organized activities after school would be beneficial for their personal and social growth. While other factors have influenced the growth of ASPs in the United States, one of the goals of many current programs is to foster youths' personal and social development through a range of adult-supervised activities. Moreover, substantial developmental research suggests that opportunities to connect with supportive adults, and participate with peers in meaningful and challenging activities in organized ASPs can help youth develop and apply new skills and personal talents (Eccles and Templeton 2002; Mahoney et al. in press; National Research Council and Institute of Medicine 2002). In other words, ASPs can be a prime community setting for enhancing young people's development.

Nevertheless, studies evaluating the personal and social benefits of ASPs have produced inconsistent findings that are further complicated by variations in the designs, participants, and types of outcomes assessed across studies (Harvard Family Research Project 2003; Mahoney et al. in press; Riggs and Greenberg 2004). Just as the meta-analysis by Lauer et al. (2006) sought to clarify the nature and extent of some of the academic benefits of ASPs, the current study applied meta-analytic techniques in an effort to examine the personal and social benefits of participation in ASPs. No previous meta-analysis has systematically examined the outcomes of ASPs that attempt to enhance youths' personal and social skills in order to describe the nature and magnitude of the gains from such programs, and to identify the features that characterize more effective programs. These are the two primary goals of the current review.

All the programs in the current review were selected because they included within their overall mission the promotion of youth's personal and social development. Although some ASPs offer a mix of activities that include academic, social, cultural, and recreational pursuits, the current review concentrates on those aspects of each program that are devoted to developing youths' personal and social skills.

## Impact of Skill Training

There is extensive evidence from a wide range of promotion, prevention, and treatment interventions that youth can learn personal and social skills (Collaborative for Academic, Social, and Emotional Learning [CASEL] 2005; Commission on Positive Youth Development 2005; Lösel and Beelman 2003). Programs that enhance children's social and emotional learning (SEL) skills cover such areas as self-awareness and self-management (e.g., self-control, self-efficacy), social awareness and social relationships (e.g., problem solving, conflict resolution, and leadership skills) and responsible decision-making (Durlak et al. 2009). Our first hypothesis was that ASPs attempting to foster participants' SEL skills would be effective and that youth would benefit in multiple ways. We examined outcomes in three general areas: feelings and attitudes, indicators of behavioral adjustment, and school performance. Positive outcomes have been obtained in these three areas for school-based SEL interventions that target youths' personal and social skills (Durlak et al. 2009), and we hypothesized that a similar pattern of findings would emerge for successful ASPs.

## Recommended Practices for Effective Skill Training

Several authors have offered recommendations regarding the procedures to be followed for effective skill training. For instance, there is broad agreement that staff are likely to be effective if they use a sequenced step-by-step training approach, emphasize active forms of learning so that youth can practice new skills, focus specific time and attention on skill training, and clearly define their goals (Arthur et al. 1998; Bond and Hauf 2004; Durlak 1997, 2003; Dusenbury and Falco 1995; Gresham 1995; Ladd and Mize 1983; Salas and Cannon-Bowers 2001). Moreover, these features are viewed as important in combination with each other rather than as independent contributing factors. For example, sequenced training will not be as effective if active forms of learning are not used, and the latter will not be as helpful unless the skills that are to be learned are clearly specified.

Although the above recommendations are drawn from skill training interventions that have primarily occurred in school and clinical settings, we expected them to be similarly important in ASPs. Therefore, we coded for the presence of the four above features using the acronym SAFE (Sequenced, Active, Focused and Explicit). We hypothesized that staff that followed all four of these features when they tried to promote personal and social skills would be more effective than staff that did not incorporate all four during skill development.

For example, new skills cannot be acquired immediately. It takes time and effort to develop new behaviors and more complicated skills must be broken down into smaller steps and sequentially mastered. Therefore, a coordinated sequence of activities is required that links the learning steps and provides youth with opportunities to connect these steps. Usually, this occurs through lesson plans or program manuals, particularly if programs use or adapt established curricula. 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).

Youth do have different learning styles, and some can learn through a variety of techniques, but evidence from many educational and psychosocial interventions indicates that the most effective and efficient teaching strategies for many youth emphasize active forms of learning. Young people often learn best by doing. Salas and Cannon-Bowers (2001) stress that "It is well documented that practice is a necessary condition for skill acquisition" (p. 480).

Active forms of learning require youth to act on the material. That is, after youth receive some basic instruction they should then have the opportunity to practice new behaviors and receive feedback on their performance. This is typically accomplished through role playing and other types of behavioral rehearsal strategies, and the cycle of practice and feedback continues until mastery is achieved. These hands-on forms of learning are much preferred over exclusively didactic instruction, which rarely translates into behavioral change (Durlak 1997).

Sufficient time and attention must be devoted to any task for learning to occur (Focus). Therefore, staff should designate time that is primarily directed at skill development. Some sources discuss this feature in terms of training being of sufficient dosage or duration. Exactly how many training sessions are needed is likely to depend on the type and nature of the targeted skills, but implicit in the notion of dosage or duration is that specific time, effort, and attention should be devoted to skills training. We coded programs on focus because of its relevance to the current meta-analysis. Although all reviewed programs indicated their intention to develop youths' personal and social skills, some did not mention any specific program components or activities that were specifically devoted to skill development. We examined how program duration related to outcomes in a separate analysis.

Finally, clear and specific learning objectives are preferred over general ones (Explicit). Youth need to know what they are expected to learn. Therefore, staff should not target personal and social development in general terms, but identify explicitly what skills in these areas youth are expected to learn (e.g., self-control, problem-solving skills, resistance skills, and so on).

In sum, the current meta-analysis of ASPs that attempt to foster the personal and social skills of program participants was conducted with the expectation that such programs would yield significant effects across a range of outcomes, and that the application of four recommended practices during the skill development components of ASPs would moderate program outcomes.

## Method

An ASP in this meta-analysis was defined as an organized program offering one or more activities that: (a) occurred during at least part of the school year; (b) happened outside of normal school hours; and (c) was supervised by adults. In addition to meeting this definition, the ASP had to meet the inclusion criterion of having as one of its goals the development of one or more personal or social skills in young people between the ages of 5 and 18. The personal and social skills could include any one or a combination of skills in areas such as problem-solving, conflict resolution, self-control, leadership, responsible decision-making, or skills related to the enhancement of self-efficacy or self-esteem. Included reports also had to have a control group, present sufficient information so that effect sizes could be calculated, and appear by December 31, 2007. Although it was not a formal criterion, all the included reports described programs conducted in the United States.

Evaluations that only focused on academic performance or school attendance and only reported academic outcomes were excluded, as were reports on adventure education and Outward Bound programs, extra-curricular school activities, and summer camps. These types of programs have been reviewed elsewhere (Bodilly and Beckett 2005; Cason and Gillis 1994; Harvard Family Research Project 2003).

### Locating Relevant Studies

The major goal of the search procedures was to secure a nonbiased representative sample of studies by conducting a systematic search for published and unpublished reports. Four primary procedures were used to locate reports: (a) computer searches of multiple databases (ERIC, PsycInfo, Medline, and *Dissertation Abstracts*) using variants of the following search terms, *after-school*, *out-of-school-time*, *school*, *students*, *social skills, youth development, children, and adolescents* (b) hand searches of the contents of three journals publishing the most outcome studies (*American Journal of Community Psychology*, *Journal of Community Psychology*, and *Journal of Counseling Psychology*), (c) inspection of the reference lists of previous ASP reviews and each included report, and (d) inspection of the

database on after-school research maintained by the Harvard Family Research Project (2009) from which many unpublished reports were identified and obtained. The dates of the literature search ranged from January 1, 1980 to December 31, 2007. Although no review can be absolutely exhaustive, we feel that the study sample is a representative group of current program evaluations.

## Study Sample

Results from 75 reports evaluating 69 different programs were evaluated. Several reports presented data on separate cohorts involved in different ASPs, each with its own control group, and these interventions were treated as separate programs. In the 75 evaluations, 68 assessed outcomes at post; 8 also collected some follow-up information, and 7 only contained follow-up data. Post effects were based on the endpoint of the youths' program participation. That is, on those occasions when two reports were available on the same participants and one contained results after 1 year of participation while the second offered information after 2 years of participation, only the latter data were evaluated. The final study sample contained examples of 21st CCLCs, programs conducted by Boys and Girls and 4-H Clubs, and a variety of local initiatives developed and supported by various community and civic organizations.

## Index of Effect

The index of effect was a standardized mean difference (SMD) that was calculated whenever possible by subtracting the mean of the control group from the mean of the after school group at post (and at follow-up if relevant) and dividing by the pooled standard deviation of the two groups. If means and standard deviations were not available, then effects were estimated using procedures described by Lipsey and Wilson (2001). When results were reported as nonsignificant and no other information was available, the effect size for that outcome measure was set at zero. There were 38 imputed zero effects and these values were not significantly associated with any coded variables.

Each effect was corrected for small sample bias and weighted by the inverse of its variance prior to any analysis (Hedges and Olkins 1985). Larger effects are desired and reflect a stronger positive impact on the after-school group compared to controls. Whenever possible, we adjusted for any pre-intervention differences between groups on each outcome measure by first calculating a pre SMD and then subtracting this pre SMD from the obtained post SMD. This strategy has been used in other meta-analyses (Derzon 2006; Wilson et al. 2001).

The consistent strategy in treating SMDs was to calculate one effect size per study for each analysis. In other words, for the first analysis of the overall effects from all 68 programs at post, we averaged all the effect sizes within each study so that each study yielded only one effect. For the subsequent analyses by outcome category, if there were multiple measures from a program for the same outcome category, they were averaged so that each study contributed only one effect size for that type of outcome. For example, if SMDs from measures of self-esteem and self-concept were available in the same study, the data were averaged to produce a single effect reflecting self-perceptions.

A random effects model was used in the analyses. A random effects model assumes that variation in SMDs across studies is the result of both sampling error and unique but random features of each study, and the use of such a model permits a broader range of generalization of the findings. A two-tailed .05 probability level was used throughout the analyses. Mean effects for different study groupings are reported along with .05 confidence intervals (CI). Moreover, homogeneity analyses were conducted to assess whether mean SMDs estimate the same population effect. Homogeneity analyses were based on the $Q$ statistic which is distributed as a chi-square with $k - 1$ degrees of freedom, where $k =$ the number of studies. For example, when studies are divided for analysis to assess possible moderator variables, $Q$ statistics assess the statistical significance of the variability in effects that exists within and between study groups. In addition, we also used the $I^2$ statistic (Higgins et al. 2003) which indicates the degree rather than the statistical significance of the variability of effects (heterogeneity) among a set of studies along a 0–100% scale.

## Coding

A coding system was developed to capture basic study features, methodological aspects of the program evaluation, and characteristics of the ASP, participants, and outcomes. The coding of most of the variables is straightforward and only a few variables are described below.

### *Methodological Features*

Two primary methodological features were coded as present or absent: use of a randomized design, and use of reliable outcome measures. The reliability of an outcome measure was considered acceptable if its alpha coefficient was $\geq$0.70, or an assessment of inter-judge agreement for coded or rated variables was $\geq$.70 (for kappa, $\geq$.60). We coded reliability in a dichotomous fashion because several reports offered no information on reliability. A third method variable, attrition, was measured on a continuous

basis as the percentage of the initial sample that was retained in the final analyses (possible range 0–100%).

*Outcome Categories*

Outcome data were grouped into eight categories. Two of these assessed feelings and attitudes (child self-perceptions and bonding to school); three were indicators of behavioral adjustment (positive social behaviors, problem behaviors, and drug use), and three assessed aspects of school performance (achievement test scores, grades, and school attendance).

*Self-perceptions*

Self-perceptions included measures of self-esteem, self-concept, self-efficacy and in a few cases (four studies) racial/cultural identity or pride. *School bonding* assessed positive feelings and attitudes toward school or teachers (e.g., liking school, or reports that the school/classroom environment or teachers are supportive). *Positive social behaviors* measured positive interactions with others. These are behavioral outcomes assessing such things as effective expression of feelings, positive interactions with others, cooperation, leadership, assertiveness in social contexts or appropriate responses to peer pressure or interpersonal conflict. *Problem behaviors* assessed difficulties that youth demonstrated in controlling their behavior adequately in social situations, and included different types of acting-out behaviors such as noncompliance, aggression, delinquent acts, disciplinary referrals, rebelliousness, and other types of conduct problems. *Drug use* primarily consisted of youth self-reports of their use of alcohol, marijuana, or tobacco. *Achievement test scores* reflected performance on standardized school achievement tests typically assessing reading or mathematics. *School grades* were either drawn from school records or reported by youth and reflected performance in specific subjects such as reading, mathematics or social studies, or overall grade point average. *School attendance* assessed the frequency with which students attended school.

SAFE Features

The presence of the four recommended practices for skill training was coded dichotomously on a yes/no basis. *Sequenced*: Does the program use a connected and coordinated set of activities to achieve their objectives relative to skill development? *Active*: Does the program use active forms of learning to help youth learn new skills? *Focused*: Does the program have at least one component devoted to developing personal or social skills? *Explicit*: Does the program target specific personal or social skills? Programs that met all four criteria were designated as SAFE programs while those not meeting all four criteria were called Other programs.

Reliability of Coding

Reliability was estimated by randomly selecting 25% of the studies that were then coded independently by the first author and trained graduate student assistants who worked at different time periods. Kappa coefficients corrected for change agreement were acceptable across all codes (0.70–0.95, average = 0.85) and disagreements in coding were resolved through discussion. The product moment correlations for coding continuous items including the calculation of effects were all above 0.95.

## Results

Table 1 summarizes several features of the 68 studies with post data. Sixty-seven per cent of the studies appeared after 2000, and the majority were unpublished technical reports or dissertations ($k = 51$, or 68%). Nearly half of the programs served elementary students (46%), over a third served students in junior high (37%), and a few involved high school students (9%; six evaluations did not report the age of participants). In terms of methodological features, 35% employed a randomized design, mean attrition was 10%, and reliability was reported and was acceptable for 73% of the outcome measures.

Twenty-five studies did not specify the ethnicity of the participants at post, and the remaining 43 reported this information in various ways. Among the latter studies, participating youth were predominantly (>90%) African American in ten studies; Latino in six studies, Asian or Pacific Islander in three studies, and American Indian in one study. There was no information on the socioeconomic status of the participants' families in nearly half of the reports ($k = 31$, or 46%). Based on the way information was reported in the remaining studies, 17 studies primarily served a low-income group (25%) and 13 studies (19%) served youth from both low- and middle-income levels.

Overall Impact at Post

First, we inspected the distribution of effects and Winsorized three values that were ≥3 standard deviations from the mean (i.e., reset these values to three standard deviations from the mean). The Winsorized study level effects, which ranged in value from −0.16 to +0.85, had an overall mean of +0.22 (CI = 0.16–0.29), which was significantly different from zero. These data indicate that ASPs have an overall positive and statistically significant impact on

**Table 1** Descriptive characteristics of reviewed studies at post

| | *k* | % |
|---|---|---|
| *Publication features* | | |
| Date of report | | |
| 1979–1990 | 3 | 4.4 |
| 1991–2000 | 19 | 27.9 |
| 2001–2008 | 46 | 67.6 |
| Source of report | | |
| Published article | 24 | 35.3 |
| Unpublished report | 44 | 64.7 |
| *Methodological features* | | |
| Experimental design | | |
| Randomized | 24 | 35.3 |
| Quasi-experimental design | 44 | 64.7 |
| Reliability of outcome measures | | |
| Acceptable reliability | 270 | 73.2 |
| Unknown/unacceptable | 99 | 26.8 |
| Mean per cent of attrition | | 10 |
| *Characteristics of participants* | | |
| Mean educational level | | |
| Elementary school (K-5) | 31 | 45.6 |
| Middle school (6–8) | 25 | 36.8 |
| High school (9–12) | 6 | 8.8 |
| Did not report | 6 | 8.8 |
| Presenting problems | | |
| None (universal intervention) | 61 | 89.7 |
| Some presenting problems | 7 | 10.3 |
| Predominant ethnicity of participants | | |
| >90% African-American | 10 | 14.5 |
| >90% Latino | 6 | 8.8 |
| >90% Asian/Pacific Islander | 3 | 4.4 |
| >90% American Indian | 1 | 1.5 |
| Did not report ethnicity | 25 | 36.8 |
| Socioeconomic status | | |
| Predominately low income | 17 | 25.0 |
| Mixed income | 13 | 19.1 |
| Did not report SES | 31 | 45.6 |
| *Program features* | | |
| Duration | | |
| Less than 1 year | 45 | 66.2 |
| 1–2 years | 12 | 17.6 |
| More than 2 years | 11 | 16.2 |

The percentages do not always add to 100% due to missing data

participating youth. However, there was statistically significant variability in the distribution of effects based on the $Q$ statistic ($Q = 306.42$, $p < .001$), and a high degree of variability according to the $I^2$ value (78%) suggesting the need to search for moderator variables that might explain this variability in program impact.

## In What Ways Do Youth Change?

Table 2 presents the mean effects obtained for the eight outcome categories, their confidence intervals, and the number of studies contributing data for each category. Significant mean effects ranged in magnitude from 0.12 (for school grades) to 0.34 for child self-perceptions (i.e., increased self-confidence and self-esteem). The mean effects for school attendance (0.10) and drug use (0.10) were the only outcomes that failed to reach statistical significance. In other words, ASPs were associated with significantly increased participants' positive feelings and attitudes about themselves and their school (child self-perceptions and school bonding), and their positive social behaviors. In addition, problem behaviors were significantly reduced. Finally, there was significant improvement in students' performance on achievement tests and in their school grades. These data support our first hypothesis. Participation in ASPs is associated with multiple benefits that pertain to youths' personal, social, and academic life.

## Moderator Analysis

There were 41 SAFE programs evaluated at post that followed all four recommended skill training practices; 27 Other programs did not use all four practices. Table 3 contains the mean SMDs for SAFE and Other Programs overall and within each of the eight outcome categories along with $Q$ and $I^2$ values. The use of $I^2$ aids in interpretation because the $Q$ statistic has low power when the number of studies is small and conversely may be statistically significant when there are a large number of studies, even though the amount of heterogeneity might be low (Higgins et al. 2003). When studies are grouped according

**Table 2** Mean effects for 68 studies at post in each outcome area

| Outcomes | SMD | *k* | 95% Confidence interval |
|---|---|---|---|
| Feelings and attitudes | | | |
| Child self-perceptions | 0.34* | 23 | {0.23, 0.46} |
| School bonding | 0.14* | 28 | {0.03, 0.25} |
| Indicators of behavioral adjustment | | | |
| Positive social behaviors | 0.19* | 36 | {0.10, 0.29} |
| Problem behaviors | 0.19* | 43 | {0.10, 0.27} |
| Drug use | 0.10 | 28 | {0.00, 0.20} |
| School performance | | | |
| Achievement test scores | 0.17* | 20 | {0.06, 0.29} |
| School grades | 0.12* | 25 | {0.01, 0.23} |
| School attendance | 0.10 | 21 | {−0.01, 0.20} |

* Denotes mean effect is significantly different from zero at the .05 level

**Table 3** Outcomes for the use of recommended skill training practices as a moderator (SAFE Criteria)

| | SAFE programs at post | | | | | Other programs at post | | | | | Between groups | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *SMD* | *k* | 95%CI | *Q*-within | $I^2$ values | *SMD* | *k* | 95%CI | *Q*-within | $I^2$ values | *Q*-between | $I^2$ values |
| All programs | 0.31* | 41 | {0.24, 0.38} | 48.07 | 17 | 0.07 | 27 | {−0.01, 0.16} | 11.94 | 0 | 17.69** | 94 |
| Feelings and attitudes | | | | | | | | | | | | |
| Child Self-perceptions | 0.37* | 21 | {0.24, 0.50} | 21.22 | 6 | 0.13 | 2 | {−0.33, 0.59} | 0.69 | 0 | 0.69 | 0 |
| School bonding | 0.25* | 13 | {0.08, 0.41} | 14.65 | 18 | 0.03 | 15 | {−0.12, 0.19} | 6.86 | 0 | 3.33 | 70 |
| Indicators of behavioral adjustment | | | | | | | | | | | | |
| Positive social behaviors | 0.29* | 19 | {0.21, 0.37} | 15.27 | 0 | 0.06 | 17 | {−0.03, 0.15} | 23.75 | 33 | 13.97** | 93 |
| Problem behaviors | 0.30* | 22 | {0.17, 0.42} | 31.74 | 34 | 0.08 | 21 | {−0.05, 0.20} | 17.01 | 0 | 5.85* | 82 |
| Drug use | 0.16* | 12 | {0.05, 0.27} | 17.86 | 38 | 0.03 | 16 | {−0.08, 0.13} | 16.36 | 8 | 2.94 | 66 |
| School performance | | | | | | | | | | | | |
| Achievement test scores | 0.20* | 10 | {0.13, 0.27} | 36.93** | 76 | 0.02 | 10 | {−0.04, 0.07} | 1.75 | 0 | 15.14** | 93 |
| Grades | 0.22* | 9 | {0.07, 0.36} | 19.71 | 60 | 0.05 | 16 | {−0.04, 0.13} | 10.71 | 0 | 3.91* | 74 |
| School attendance | 0.14** | 9 | {0.05, 0.24} | 10.69 | 25 | 0.07 | 12 | {0.01, 0.13} | 10.52 | 0 | 1.65 | 39 |

* Denotes mean effect is significantly different from zero at the .05 level

** Denotes mean effect is significantly different from zero at the .01 level

to hypothesized moderators, there should be low heterogeneity *within* groups (reflected in low $I^2$ values and non-significant *Q* statistics) but high and statistically significant levels of heterogeneity *between* groups (reflected by corresponding high $I^2$ values and statistically significant *Q*-between values). Benchmarks for $I^2$ suggest that values under 15% indicate negligible heterogeneity, from 15 to 24% reflect a mild degree of heterogeneity, between 25 and 50% a moderate degree, and values ≥75% a high degree of heterogeneity (Higgins et al. 2003).

The data in Table 3 indicate that whereas SAFE programs are associated with significant mean effects for all outcomes (mean SMDs between 0.14 and 0.37), Other programs do not yield significant mean effects for any outcome. There is empirical support for moderation for four outcomes in terms of significant *Q*-between statistics and correspondingly high (74–93%) $I^2$ values (positive social behaviors, problem behaviors, achievement test scores, and grades). However, the *Q*-between statistics were not significant and the $I^2$ values were generally low for the other four outcomes (self-perceptions, school bonding, drug use and school attendance). Furthermore, there is a moderate degree of within group variability among SAFE programs ($I^2$ values between 34 and 76%) for four outcomes (problems behaviors, drug use, test scores and grades) suggesting the possibility of additional moderators that might improve the model fit.

Although we required that program staff had to follow all four SAFE practices, there was some relationship between the absolute number of practices used and outcomes. The mean study-level ESs for staff using none, one, two, or four of the SAFE practices (three practices were not present in any report) were 0.02 ($k = 4$), 0.07 ($k = 7$), 0.10 ($k = 16$), and 0.31 ($k = 0.31$), respectively.

## Ruling out Rival Explanations

To examine other potential explanations for the results we first compared the effects in each outcome category for studies grouped according to each of the following variables: randomization (yes or no), use of a reliable outcome measure (yes or no), presence of an academic component in the ASP (yes or no), and the educational level (elementary, middle, or high school) and gender of the participants. We also computed product moment correlations between SMDs and sample size, program duration, and per cent of attrition. There were too few data on participants' ethnicity and socioeconomic status to examine these variables adequately. Setting was strongly associated with the presence of an academic component so we only examined the latter variable (i.e., school-based programs were more likely to offer some form of academic assistance).These procedures resulted in 64 analyses (eight variables crossed with eight outcome categories). For these analyses, significant effects emerged in only two cases, which would be expected by chance. The use of randomized designs was associated with higher levels of positive social behaviors (*Q*-between = 4.80, $p < .05$), and there was a significant positive

correlation between female gender and higher test scores ($r = 0.69, p < .01$). Overall, these analyses suggest that the above variables do not serve as an alternative explanation for the positive findings obtained by SAFE programs. Additional comparisons indicated that SAFE and Other programs did not differ significantly on any of the above variables.

We also examined publication source and found that published reports ($k = 24$) yielded significantly higher study-level ESs than unpublished ($k = 44$) reports (respective mean SMDs = 0.34 and 0.10). Upon further examination, this effect was restricted to Other programs. Whereas 25 unpublished Other studies yielded a nonsignificant mean SMD of 0.05 (CI = −0.03, 0.13), the two published studies of Other programs had a mean SMD of 0.69 (CI = 0.21, 1.17). In contrast, there was no SMD mean difference between the 19 unpublished and 22 published reports of SAFE programs (mean SMDs = 0.31 and 0.30, respectively).

## Sensitivity Analyses

We conducted several sensitivity analyses of study-level effects in consideration of different features of program evaluations. Among the 68 evaluations at post, eight were conducted by researchers who had apparently developed the skills content of the ASP and might have a vested interest in its positive outcomes; there were 17 studies in which investigators failed to confirm the pre-intervention equivalence of program and control groups; there were four cases of differential attrition occurring between program and control groups; and in eight cases, some criterion related to attendance was used when composing the intervention sample (e.g., only children who attended a certain percentage of available times were assessed). On the other hand, in 22 studies an intent-to-treat analysis was conducted in which all youth assigned to the ASP were assessed regardless of whether or not they attended frequently or not at all. In two cases, the same ASP was evaluated in two separate reports. Separate analyses removing the studies with each of above features did not change the main outcome findings.

Finally, although published and unpublished SAFE studies yielded similar results, we also conducted a trim and fill analysis (Duval and Tweedie 2000) to estimate the possibility of publication bias on study-level effect sizes (i.e., to determine if additional but missing unpublished studies would change the main finding). This procedure suggested the trimming and filling of four studies and resulted in an adjusted mean estimate for SAFE programs that remained statistically significant from zero (mean ES = 0.22, $p < .05$).

## The Impact of Pre SMDs

The impact of computing pre SMDs is reflected in the mean comparisons between the 81 outcomes in which it was possible to calculate such SMDs (group 1) versus the remaining 334 outcomes in which these data could not be calculated due to lack of information (group 2). While the post mean SMDs are similar for both groups (0.20 and 0.18, respectively), the mean pre SMD for group 1 was −0.10. Subtracting the pre SMD from the post SMD to create an adjusted post SMD for group 1 produced a significant mean difference at post favoring group 1 (0.29 versus 0.18, respectively, $p < .01$). The values of the pre and post mean SMDs for group 1 indicate that on 20% of the outcomes the after-school group started at a documented disadvantage compared to controls, but overcome this disadvantage over time and were superior to the control group at post. Including pre SMDs increased the overall mean effect by 61% on these outcomes. (0.29 vs. 0.18). Pre SMDs were not more likely for some outcome categories than others, nor were they associated with other coded variables except for SAFE and Other programs. The former were more likely to have pre SMD which might be one factor contributing to their larger effects.

## Putting Current Findings into Context

It may seem customary to view the effects achieved in this review (i.e., mean SMDs in the 0.20 and 0.30s) as "small" in magnitude. However, methodologists now stress that instead of simply resorting to Cohen's (1988) conventions regarding the size of obtained effects, findings should be interpreted in the context of prior research and, whenever possible, in terms of their practical value (Vacha-Haase and Thompson 2004). If one does so, the impact for ASP programs achieves more prominence.

For example, Table 4 compares the mean SMDs achieved by the 41 effective SAFE programs to the results reported in meta-analyses of other interventions for school-aged youth. The SMDs of SAFE programs are similar to or better than those produced by several other community- and school-based interventions for youth assessing outcomes such as self-perceptions, positive social behaviors, problem behaviors, drug use, and school performance (DuBois et al. 2002; Durlak and Wells 1997; Haney and Durlak 1998; Lösel and Beelman 2003; Tobler et al. 2000; Wilson et al. 2001, 2003). For these comparisons, we used the findings from other meta-analyses regarding universal interventions wherever possible because the vast majority of effective ASPs in our review did not involve youth with identified problems.

Of particular note, the mean SMD obtained by SAFE programs on achievement test scores (0.31) is not only

**Table 4** Comparing the mean effects of SAFE programs to the results of other universal interventions for children and adolescents

| Outcomes | Current review | Other reviews |
|---|---|---|
| *Mean effects* | | |
| Feelings and attitudes | | |
| Self-perceptions | 0.37 | 0.19[a] |
| School bonding | 0.25 | – |
| Indicators of behavioral adjustment | | |
| Positive social behaviors | 0.29 | 0.15[b], 0.39[c] |
| Problem behaviors | 0.30 | 0.21[b], 0.27[c], 0.09[d], 0.17[e] 0.30[f] |
| Drug use | 0.16 | 0.11[b], 0.05[e], 0.15[g] |
| School performance | | |
| Achievement test scores | 0.20 | 0.11[b], 0.30[f], 0.24[h] |
| Grades | 0.22 | – |
| School attendance | 0.14 | – |

Results from other meta-analyses are from outcome categories most comparable to those in the current review and resulting from weighted random effects analyses whenever possible

[a] Haney and Durlak (1998), [b] DuBois et al. (2002), [c] Lösel and Beelman (2003), [d] Wilson et al. (2003), [e] Wilson et al. (2001), [f] Durlak and Wells (1997), [g] Tobler et al. (2000), [h] Hill et al. (2008)

larger than the effects obtained in reviews of primarily academically-oriented ASPs and summer school programs (Cooper et al. 2000; Lauer et al. 2006), but is comparable to the results of 87 meta-analyses of school-based educational interventions (Hill et al. 2008).

It is possible to convert a mean SMD into a percentile using Cohen's $U_3$ index to reflect the average difference between the percentile rank of intervention and control groups (Institute for Education Sciences 2008a). A mean effect of 0.31 translates into a percentile difference of 12%. Put another way, the average member of the control group would demonstrate a 12 percentile increase in achievement if they had participated in a SAFE after-school program.

### Results at Follow-up

The 15 reports containing follow-up data collected information on different outcome categories. The cell sizes at follow-up ranged from zero for school attendance to nine for self-perceptions (mean ES = 0.19; $p < .05$). Unfortunately, there is too little information at follow-up to offer any conclusions about the durability of changes produced by ASPs.

## Discussion

This is the first meta-analysis to evaluate the outcomes achieved by ASPs that seek to promote youths' personal and social skills. This review included a large number of ASPs ($k = 75$), and is the first time many of these reports have been scrutinized. Two-thirds of the evaluated reports appeared after 2000. As a result, this review yields an up-to-date perspective on a rapidly growing research literature.

Current data indicate that ASPs had an overall positive and statistically significant impact on participating youth. Desirable changes occurred in three areas: feelings and attitudes, indicators of behavioral adjustment, and school performance. More specifically, there were significant increases in youths' self-perceptions, bonding to school, positive social behaviors, school grades, and achievement test scores. Significant reductions also appeared for problem behaviors. Finally, SAFE programs were associated with practical gains in participants' test scores suggesting an average difference of 12 percentile points between the after-school and control group, and achieved results in this and several other areas that were similar to or better than those obtained by many other evidence-based psychosocial interventions for school-aged populations. The implication of current findings is that ASPs merit support and recognition as an important community setting for promoting youths' personal and social well-being and adjustment.

An important qualification is that not all ASPs were effective. Only the group of SAFE programs yielded significant effects on any outcomes. Commenting on the results of our review as well as several others, Granger (2008) noted that although some ASPs achieve positive results, many others do not, indicating that there is much room for improvement among current programs. As we discuss, below, this has important implications for future research and practice.

Several steps were taken to increase the credibility of the findings. We searched carefully and systematically for relevant reports to obtain a representative sample of

published and unpublished evaluations, and are confident that our sample of studies is an unbiased representation of evaluations of ASPs meeting our inclusion criteria that have appeared by the end of 2007. We also examined and were able to rule out some plausible rival explanations for our main findings. Furthermore, the current review underestimates the true impact of ASPs for at least two reasons. One has to do with the nature of the control groups used in current evaluations; the second has to do with the dosage of the intervention received by many program youth.

### Control Groups

The intent of this review was to compare outcomes for youth attending a particular ASP to those not attending the program, but this does not mean that comparison youth constituted a true no intervention control group. For example, it is well known that in any one time period not only do many youth spend their out-of-school time in different pursuits (e.g., in ASPs, extra-curricular school activities and church groups, as well as hanging out with friends, and being alone some of the time), but also they may change their level of participation across activities over time (Mahoney et al. 2006, in press). In five reviewed reports, authors noted that youth in their control condition were participating in alternative ASPs or other types of potentially beneficial out-of-school time activities (Brooks et al. 1995; Philliber et al. 2001; Rusche et al. 1999; Tebes et al. 2007; Weisman et al. 2003). It is recommended that evaluators monitor the types of alternative services that are received by control groups, so a truer estimate of the impact of intervention can be made.

### Program Dosage

It is axiomatic that recipients must receive a sufficient dosage for an intervention to have an effect. However, it appears this did not happen in several of the reviewed programs, which may be an explanation for the poor results obtained for some programs. Although each report did not contain specific data on program attendance, when some information was presented it was apparent that attendance was a problem for several programs. For example, youths' attendance ranged from 15 to 26% in 11 evaluations (Baker and Witt 1996; Dynarski et al. 2004; James-Burdumy et al. 2005; LaFrance et al. 2001; Lauver 2002; Maxfield et al. 2003, two cohorts; Philliber et al. 2001; Prenovost 2001, three cohorts).

Moreover, analyses conducted in some reports indicated that attendance was positively related to youth outcomes. This occurred in six of the seven studies that examined this issue, although significant differences did not always emerge on every outcome measure (Baker and Witt 1996; Fabiano et al. 2005; Lauver 2002; Morrison et al. 2000; Prenovost 2001; Vandell, et al. 2005; Zief 2005). Reviews of other ASPs have also reported a significant positive relationship between attendance and positive outcomes (Simpkins et al. 2004, but also see Roth et al. 2010).

Furthermore, attendance is only one aspect of participation. Information is also needed on the breadth of youth activities within any program and their level of engagement in each activity. For example, studies suggest that youths' level of engagement predicts positive social and academic outcomes (Mahoney et al. 2007; Shernoff 2010). In sum, the receipt of alternative after-school activities by control groups and the low attendance achieved in some programs worked against finding positive outcomes. The next sections discuss several other issues suggested by the current findings.

### Elements of Effective ASPs

As hypothesized, the use of four recommended training practices (i.e., SAFE) moderated several outcomes and distinguished between ASPs that were or were not associated with multiple positive outcomes. Moreover, there is convergent evidence from numerous other sources on the importance of SAFE features. Although the terminology may differ, others have mentioned the importance of one or more SAFE features in ASP programs (Gerstenblith et al. 2005; Granger and Kane 2004; Mahoney et al. 2001, 2002; Miller 2003, National Research Council and Institute of Medicine 2002). For example, Granger (2008) noted that our data were consistent with a developing consensus in the after-school field that "being explicit about program goals, implementing activities focused on these goals, and getting youth actively involved are practices of effective programs" (p. 11). We recommend that future research should continue to examine the value of these features in ASPs. Fortunately, SAFE practices can be applied to a wide variety of intervention approaches.

### Gains in Achievement Test Scores

SAFE ASPs yielded significant improvement in participants' standardized test scores and at a magnitude (i.e., SMD of 0.31), which is over two times larger than that found in the previous meta-analysis of academically-oriented ASPs (Lauer et al. 2006). Why were current programs so effective in the academic realm?

There are several possible explanations. First, it should come as no surprise that programs promoting skill development can also improve school performance. There is now a growing body of research indicating that interventions that promote SEL skills also result in improved academic performance (Collaborative for Academic, Social,

and Emotional Learning [CASEL] 2005; Weissberg and Greenberg 1998; Zins et al. 2004). We have obtained a mean SMD of similar magnitude (i.e., 0.27) for school-based interventions promoting students' personal and social skills (Durlak et al. 2009).

Second, current results are based on a set of recent evaluations of ASPs, only a few of which have ever been part of any previous review. Although we did not code the academic components of ASPs, it is possible that developers of newer ASPs may have used strategies that would strengthen their impact. For example, others have suggested that gains in academic achievement are more likely to occur if staff are well-trained and supervised, use evidence-based instructional strategies, are supportive and reinforcing to youth during learning activities, conduct pre-assessments to ascertain learners' strengths and academic needs, and coordinate their teaching or tutoring with school curricula (e.g., Birmingham et al. 2005; Southwest Educational Development Laboratory 2006). A recent multi-site evaluation indicated that ASP participants do manifest academic progress if evidence-based instructional strategies are used and are well-implemented (Sheldon et al. in press). More research needs to analyze how different features of the academic components of future ASPs contribute to outcomes. Third, it must be acknowledged that only 20 programs collected outcome data on academic achievement, so current results need replication in more programs to confirm their generality.

## Limitations and Directions for Future Research

There are four important limitations in our review that suggest directions for future research.

1. Current conclusions rest upon outcome research that should be improved in several ways. Many reports lacked data on the racial and ethnic composition or the socio-economic status of participants, so we could not relate outcomes to these participant characteristics. Missing statistical data at pre or post limited the number of effects that could be directly calculated. At a minimum, future program evaluations should provide complete information on the demographic characteristics of participants, their pre and post scores on all outcomes, and, if pertinent, their prior academic achievement, and any presenting problems youth might have. The goals, procedures and contents of each program component should be specified and described, and data on levels of participation and breadth and degree of engagement in different activities should be included. Reliable and valid outcome measures should be used and, whenever possible, data should be collected using multiple methodologies (e.g., from school records, questionnaires, and behavioral observations) and from multiple informants (e.g., youth, parents, teachers, and ASP staff).

Future evaluations should also be aware of the analytic procedures that should be used for nested designs. That is, when an intervention is conducted in a group context or setting such as in an ASP, participant data are not independent and analyses treating individual data as independent can greatly increase Type I error rates. Unfortunately, virtually all the reviewed reports employed one intervention and one control group so that appropriate corrections for nested data could not be made (Baldwin et al. 2005). Guidelines are available for the appropriate analyses of nested data (Institute for Education Sciences 2008b; Raudenbush and Bryk 2002).

Care is also needed in designating program participants. Eight studies only analyzed data from participants who had attended a certain number of program activities using unique criteria in each circumstance. This method confounds the impact of intervention with dosage. A preferred strategy used in some studies (e.g., Philliber et al. 2001; Weisman et al. 2001; Zief 2005) is an intent-to-treat analysis in which all participants' data are evaluated regardless of their program dosage. Additional analyses can then be conducted to examine the relationship between program attendance and outcomes.

Current findings illustrate how the impact of intervention can be more completely portrayed by including pre SMDs in the final calculation of effects. On 19% of the outcomes, the after-school group started at a disadvantage (mean pre SMD $= -0.10$) but overcame this disadvantage over time (mean post SMD $= 0.20$). Incorporating pre SMDs increased the final SMD for these outcomes by 69% (0.29 versus 0.20). More journals are now requiring authors to report SMDs for individual studies (Durlak 2009) and future researchers should consider calculating adjusted SMDs that take into consideration any initial differences between groups.

2. Although the four SAFE features we assessed did distinguish between more and less effective programs, it is important to put these findings in context. First, authors have noted additional aspects of skill training that are important, such as the trainer's interpersonal skills, sensitivity to the learner's developmental abilities and cultural background, and the importance of helping youth generalize their newly-developed skills to everyday settings (Dusenbury and Falco 1995; Gresham 1995). Unfortunately, information on these additional recommended elements was not available.

Second, although previous authors have stressed that the four features we assessed work in combination with each other, their relative influence might nevertheless vary not only in relation to youths' developmental level and cultural background, but also on the nature and number of targeted skills. For example, younger children will likely need more practice than older youth when attempting to master more

complex skills. The relative influence of different training procedures on eventual skill development also deserves attention in future research.

Third, 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 use of active learning techniques instead of viewing these practices as all-or-none phenomena. Observational systems have now been developed to record the use of SAFE practices in ASPs as continuous variables (Pechman et al. 2008).

Fourth, based on $Q$ and $I^2$ values there was stronger empirical support for SAFE practices as moderators for some outcomes over others (e.g., for positive social behaviors, problem behaviors, test scores, and grades) and it was possible to calculate pre SMDs for more SAFE than Other programs. Therefore, current data are just a beginning in exploring the "black box" of ASPs, that is, in understanding all the structures and processes that constitute an effective program. Current data are correlational in nature and we cannot conclude that SAFE features caused the positive changes in program participants. Because the current meta-analysis focused only on the skill-building components of ASPs, it is possible that additional program variables play a role in the effectiveness of ASPs. For example, program quality is one feature that comes to mind, and has been emphasized in the operation of ASPs (Birmingham et al. 2005; Granger 2008; High/Scope Educational Research Foundation 2005; Miller 2003; Vandell et al. 2004).

Several independent groups have focused on six core features that contribute to the quality of a youth development program (Yohalem et al. 2007). In addition to skill-building opportunities, these include the characteristics of interpersonal relationships (between staff and youth and among youth), the program's psychosocial environment, the level and nature of youths' engagement in activities, social norms, and program routine and structure. In turn, these features are related to such variables as staff behavior, program policies, youth initiative, and issues related to community partnerships and support. Information on these variables were not available in reviewed reports, and future researchers should explore their influence.

Two additional, important foci for future research involve the creation and evaluation of effective staff development and training programs and data on program implementation. What are the most efficient ways for staff to learn new techniques and implement them effectively?

Some authors stress the importance of less structured activities that might stimulate youth initiative and foster heightened leadership skills and autonomy. The South Baltimore Youth Center (Baker et al. 1995) which did not follow SAFE practices used an empowering strategy by having participants assume responsibility for all major Youth Center activities. This strategy was associated with impressive significant improvement in adolescents' self-reported delinquent behavior and drug use (SMDs of 1.10 and 0.82, respectively). Data from other studies also confirm the value of empowering strategies in ASPs (Hirsch and Wong 2005; Hansen and Larson 2007), but more controlled outcome data are needed.

Nevertheless, the findings on the value of structured skill development practices do not necessarily contradict the value of less structured activities for three reasons. First, alternative strategies can lead to similar outcomes. There are many possible ways to get from Point A to Point B and some competencies may be better promoted via one strategy than another. Studies directly comparing the relative benefits of different strategies on different skills and adjustment outcomes would be helpful. Second, most ASPs contain multiple components so more structured approaches can be used some of the time and less structured ones at other times. Third, empowerment strategies can be used within structured components, for example, by asking more skilled youth to be role models, trainers, or co-group leaders for others. Assuming such roles could promote youths' leadership skills and sense of self-efficacy.

Future research that can clarify how different aspects of program quality influence different youth outcomes will be extremely helpful in improving ASPs. Because program quality is a multi-dimensional construct, assessing quality across its dimensions and relating these to a range of youth outcomes can provide an empirical basis for understanding the processes within ASPs that lead to different results. As research on this topic accumulates, it will be possible to develop a clearer understanding of what constitutes a high quality program and in what respects current programs can be improved.

3. Unfortunately, few reports have collected any follow-up data, so we cannot offer any conclusions about the long-term effects of ASPs. Hopefully, future evaluations will contain follow-up information to determine the durability of any gains emanating from program participation.

4. Although the initial study sample seems sufficiently large (68 studies with post data), dividing studies first according to outcome categories, and then according to other potentially important variables reduced the statistical power of the analyses. Therefore, the failure to obtain statistically significant findings for some of the variables examined here should be viewed cautiously.

As more ASP evaluations appear, researchers will have more power to detect the influence of potentially important variables. At the individual level, we need information on how gender, race/ethnicity, age, income status, and the presence of academic or behavioral problems are related to participants' participation, engagement and different types

of outcomes. At the ecological level we need to understand how family, school, and neighborhood characteristics and resources are associated with consistent and active participation in ASPs, and interact with various program processes and structures to influence youth outcomes (Mahoney et al. 2007; Weiss et al. 2005). Such data would help us maximize the fit between program features and local needs to increase the reach and benefits of ASPs.

Notwithstanding the above limitations, the current review offers empirical support for the notion that ASPs can be successful in achieving their historical mission of fostering the personal and social development of young people. Although not conclusive, current findings should stimulate more interest in investigating and understanding how ASPs programs affect youth, and what can be done to enhance their effectiveness.

**Acknowledgments** This article is based on a grant from the William T. Grant Foundation (grant #2212) awarded to the first and second authors. We wish to express our appreciation to David DuBois, Mark Lipsey, Robert Granger, and Nicole Yohalem who provided helpful comments on an earlier draft of this manuscript. We offer additional thanks to Mark Lipsey and David Wilson for providing the macros used for calculating effects from each relevant outcome and conducting the statistical analyses. Finally, we wish to thank Heather Weiss and Chris Wimer from the Harvard Family Research Project who supplied copies of relevant reports that we were unable to obtain.

## References

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

Arthur, W., Jr., Bennett, W., Jr., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11*, 57–101.

*Astroth, K. A., & Haynes, G. W. (2002). More than cows and cooking: Newest research shows the impact of 4-H. *Journal of Extension, 40*, 1–10.

*Baker, K., Pollack, M., & Kohn, I. (1995). Violence prevention through informal socialization: An evaluation of the South Baltimore Youth Center. *Studies on Crime and Prevention, 4,* 61–85.

*Baker, D., & Witt, P. A. (1996). Evaluation of the impact of two after-school programs. *Journal of Park and Recreation Administration, 14*, 60–81.

Baldwin, S. A., Murray, D. M., & Shadish, W. R. (2005). Empirically supported treatments or Type I errors? Problems with the analysis of data from group administered treatments. *Journal of Consulting and Clinical Psychology, 73*, 924–935.

*Belgrave, F. Z., Chase-Vaughn, G., Gray, F., Addison, J. D., & Cherry, V. R. (2000). The effectiveness of a culture- and gender-specific intervention for increasing resiliency among African American preadolescent females. *Journal of Black Psychology, 26,* 133–147.

*Bergin, D. A., Hudson, L. M., Chryst, C. F., & Resetar, M. (1992). An afterschool intervention program for educationally disadvantaged young children. *The Urban Review, 24,* 203–217.

Birmingham, J., Pechman, E. M., Russell, C. A., & Mielke, M. (2005). *Shared features of high-performing after-school programs: A follow-up to the TASC evaluation*. Washington, DC: Policy Studies Associates. Retrieved May 29, 2007 from www.sedl.org/pubs/catalog/items/fam107.html.

*Bissell, J., Dugan, C., Ford-Johnson, A., Jones, P., & Ashurst, J. (2002). *Evaluation of the YS-CARE after school program for California work opportunity and responsibility to kids (CalWORKS)*. Department of Education, University of California, Irvine and Research Support Services.

Bodilly, S., & Beckett, M. K. (2005). *Making out-of-school time matter: Evidence for an action agenda.* Santa Monica, CA: Rand Corporation. Retrieved September 10, 2005, from www.rand.og//pubs/monographs/MG242/index.html.

Bond, L. A., & Hauf, A. M. C. (2004). Taking stock and putting stock in primary prevention: Characteristics of effective programs. *Journal of Primary Prevention, 24*, 199–221.

*Brooks, P. E., Mojica, C. M., & Land, R. E. (1995). *Final evaluation report: Longitudinal study of LA's BEST after school education and enrichment program, 1992–1994*. Los Angeles: University of California, Graduate School of Education & Information Studies, Center for the Study of Evaluation.

Cason, D., & Gillis, H. L. L. (1994). A meta-analysis of outdoor adventure programming with adolescents. *Journal of Experiential Education, 17*, 40–47.

*Chase, R. A. (2000). *Hmong American partnership: 2HTN final report*. St. Paul, MN: Wilder Research Center.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hilllsdale, NJ: Erlbaum.

Collaborative for Academic, Social, and Emotional Learning [CASEL]. (2005). *Safe and sound: An educational leader's guide to evidence-based social and emotional learning programs—Illinois edition.* Retrieved January 10, 2007, from http://www.casel.org.

Commission on Positive Youth Development. (2005). The positive perspective on youth development. In D. W. Evans, E. B. Foa, R. E. Gur, H. Hendin, C. P. O'brien, M. E. P. Seligman, & B. T. Walsh (Eds.), *Treating and preventing adolescent mental health disorders: What we know and what we don't know* (pp. 497–527). NY: Oxford University Press.

Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development, 65*(1), 260.

Derzon, J. (2006). How effective are school-based violence prevention programs in preventing and reducing violence and other antisocial behaviors? A meta-analysis. In S. R. Jimerson & J. J. Furlong (Eds.), *The handbook of school violence and school safety: From research to practice* (pp. 429–441). Mahwah, NJ: Lawrence Earlbaum.

DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology, 30*, 157–198.

Durlak, J. A. (1997). *Successful prevention programs for children and adolescents*. New York: Plenum.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34*, 917–928.

Durlak, J. A.,Weissberg, R. P., Dymnicki, A. B., Taylor, R, D., & Schellinger, K. B. (2009). The impact of enhancing students' social and emotional development: A meta-analysis of school-based universal interventions. Manuscript submitted for publication.

Durlak, J. A., & Wells, A. M. (1997). Primary prevention mental health programs for children and adolescents: A meta-analytic review. *American Journal of Community Psychology, 25*, 115–152.

Dusenbury, L., & Falco, M. (1995). Eleven components of effective drug prevention curricula. *Journal of School Health, 65*, 420–425.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455–463.

*Dynarski, M., James-Burdumy, S., Moore, M., Rosenberg, L., Deke, J., & Mansfield, W. (2004). *When schools stay open late: The national evaluation of the 21st Century Community Learning Centers Program: New findings*. US Department of Education, National Center for Education Evaluation and Regional Assistance. Washington, D.C.: US Government Printing Office.

Eccles, J. S., & Templeton, J. (2002). Extracurricular and other after-school activities for youth. *Review of Research in Education, 26*, 113–180.

*Fabiano, L., Pearson, L. M., & Williams, I. J. (2005). *Putting students on a pathway to academic and social success: Phase III findings of the Citizen Schools evaluation*. Washington, DC: Policy Studies Associates, Inc.

*Foley, E. M., & Eddins, G. (2001). *Preliminary analysis of Virtual Y after-school program participants' patterns of school attendance and academic performance: Final evaluation report program year 1999–2000*. NY: National Center for Schools and Communities, Fordham University.

*Fuentes, E. G. (1983). A primary prevention program for psychological and cultural identity enhancement: Puerto Rican children in semi-rural northeast United States. *Dissertation Abstracts International, 44* (05), 1578B.

Gerstenblith, S. A., Soule, D. A., Gottfredson, D. C., Lu, S., Kellstrom, M. A., & Womer, S. C. (2005). After-school programs, antisocial behavior, and positive youth development: An exploration of the relationship between program implementation and changes in youth behavior. In J. L. Mahoney, J. S. Eccles, R. W. Larson, et al. (Eds.), *Organized activities as contexts of development: Extracurricular activities, after-school and community programs* (pp. 457–478). Mahwah, NJ: Erlbaum.

*Gottfredson, D. C., Soule, D. A., & Cross, A. (2004). *A statewide evaluation of the Maryland after school opportunity fund program*. Department of Criminology and Criminal Justice, University of Maryland.

Granger, R. C. (2008). After-school programs and academics: Implications for policy practice, and research. *Social Policy Report, 22*(3–11), 14–19.

Granger, R. C., & Kane, T. (2004). Improving the quality of after-school programs. *Education Week, 23*, 76–77.

*Grenawalt, A., Halback, T., Miller, M., Mitchell, A., O'Roarke, B., Schmitz, T., et al. (2005). *4-H animal science program evaluation: Spring 2004–What is the value of the Wisconsin 4-H Animal Science Projects?* Madison WI: University of Wisconsin Cooperative Extension.

Gresham, F. M. (1995). Best practices in social skills training. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-III* (pp. 1021–1030). Washington, DC: National Association of School Psychologists.

*Hahn, A., Leavitte, T., & Aaron, P. (1994). *Evaluation of the Quantum Opportunities Program (QOP): Did the program work?* Waltham, MA: Brandeis University, Heller Graduate School, Center for Human Resources.

Haney, P., & Durlak, J. A. (1998). Changing self-esteem in children and adolescents: A meta-analytic review. *Journal of Clinical Child Psychology, 27*, 423–433.

Hansen, D. M., & Larson, R. W. (2007). Amplifiers of developmental and negative experiences in organized activities: Dosage, motivation, lead roles, and adult-youth ratios. *Journal of Applied Developmental Psychology, 28*, 360–374.

Harvard Family Research Project. (2003). *A review of out-of-school time program quasi-experimental and experimental evaluation results*. Cambirdge, MA: Harvard Family Research Project.

Harvard Family Research Project (2009). *Out-of-school-time program research and evaluation database and bibliography*. Retrieved June 5, 2009 from http://www.hfrp.org/out-of-school-time/ost-database-bibliography.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. NY: Academic Press.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557–560.

High/Scope Educational Research Foundation. (2005). *Youth program quality assessment validation study: Findings for instrument validation.* Retrieved April 6, 2006 from http://www.highschope.org/EducationalPrograms/Adolescent/YouthPQA/YouthPQASummary.pdj.

Hill, J. C., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*, 172–177.

Hirsch, B., & Wong, V. (2005). *A place to call home: After-school programs for urban youth*. Washington, DC: American Psychological Association and New York: Teachers College Press.

*Huang, D. (2004). *Exploring the long-term impact of LA's Best on student's social and academic development*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

*Huang, D., Sung Kim, K., Marshall, A., & Perez, P. (2005). Keeping kids in school: An LA's BEST example National Center for Research on Evaluation, Standards, and Student Testing (CRESST) Center, Center for the Study of Evaluation (CSE), Graduate School of Education and Information Studies, University of California, Los Angeles.

*Hudley, C. (Ed.). (1999). *Problem behaviors in middle childhood: Understanding risk status and protective factors*. Montreal, Quebec, Canada: California Wellness Foundation. (ERIC Document Reproduction Service No. ED 430 066).

Institute for Education Sciences. (2008a). What works clearinghouse procedures and Standards Workbook, Version 2.0, December, 2008. Retrieved June 18, 2009 from http/ies.ed.gov/ncee/wwc/references/idocViewer/Doc.aspx?docid=198tocid=1.

Institute for Education Sciences. (2008b). *Technical details of WWC-conducted computations*. Retrieved June 6, 2008 from http//ies.ed.gov.ncee.wwc/pdf/conducted_computations.pdf.

*James-Burdumy, S., Dynarski, M., Moore, M., Deke, J., & Mansfield, W. (2005). When school stay open late: The national evaluation of the 21st Century Community Learning Centers Program Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Kane, T. J. (2004). *The impact of after-school programs: Interpreting the results of four recent evaluations*. Retrieved January 17, 2006 from www.wtgrantfondation.org/usr_doc/After-school_paper.pdf.

Ladd, G. W., & Mize, J. (1983). A cognitive social learning model of social skill training. *Psychological Review, 90*, 127–157.

*LaFrance, S., Twersky, F., Latham, N., Foley, E., Bott, C., Lee, L., et al. (2001). *A safe place for healthy youth development: A comprehensive evaluation of the Bayview Safe Haven*. San Francisco, CA: BTW Consultants and LaFrance Associates.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Green, M. (2006). Out-of school time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research, 76*, 275–313.

*Lauver, S. C. (2002). Assessing the benefits of an after-school program for urban youth: An impact and process evaluation. *Dissertation Abstracts International*, 63 (02), 533A.

*LeCroy, W. W. (2004). Experimental evaluation of "Go Grrrls" preventive intervention for early adolescent girls. *Journal of Primary Prevention, 25,* 457–473.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

*LoSciuto, L., Hilbert, S. M., Fox, M. M., Porcellini, L., & Lanphear, A. (1999). A two-year evaluation of the Woodrock Youth Development Project. *Journal of Early Adolescence, 19,* 488–507.

Lösel, F., & Beelman, A. (2003). Effects of child skills training in preventing antisocial behavior: A systematic review of randomized evaluations. *Annals of the American Academy of Political and Social Science, 587*, 84–109.

Mahoney, J. P., Parente, M. E., & Zigler, E. F. Afterschool program participation and children's development. In J. Meece & J. Eccles (Eds.), *Handbook of research on schools, schooling, and human development.* New York: Wiley (in press).

*Mahoney, J. L., Lord, H., & Carryl, E. (2005). Afterschool program participation and the development of child obesity and peer acceptance. *Applied Developmental Science, 9*, 202–215.

Mahoney, J. L., Parente, M. E., & Lord, H. (2007). Program-level differences in afterschool program engagement: Links to child competence, program quality and content. *The Elementary School Journal, 107*, 385–404.

Mahoney, J. L., Schweder, A. E., & Stattin, H. (2002). Structured after-school activities as a moderator of depressed mood for adolescents with detached relations to their parents. *Journal of Community Psychology, 30*, 69–86.

Mahoney, J. L., Stattin, H., & Magnusson, D. (2001). Youth recreation centre participation and criminal offending: A 20-year longitudinal study of Swedish boys. *International Journal of Behavioral Development, 25*, 509–520.

Mahoney, J. P., & Zigler, E. F. (2006). Translating science to policy under the No Child Left Behind act of 2001: Lessons from the national evaluation of the 21st-Century Community Learning Centers. *Journal of Applied Developmental Psychology, 27*, 282–294.

*Mason, M. J., & Chuang, S. (2001). Culturally-based after-school art programming for low-income urban children: Adaptive and preventive effects. *Journal of Primary Prevention, 22,* 45–54.

*Maxfield, M., Schirm, A., & Rodriguez-Planas, N. (2003). *The Quantum Opportunity Program demonstration: Implementation and short-term impacts*. Washington, DC: Mathematica Policy Research.

*McClanahan, W. S., Sipe, C. L., & Smith, T. J. (2004). *Enriching summer work: An evaluation of the summer career exploration program*. Philadelphia, PA: Public/Private Ventures.

Miller, B. M. (2003). *Critical hours: After-school programs and educational success.* New York: Nellie Mae Education Foundation. Retrieved December 13, 2005 from www.nmefdn.org/uploads/Critical_hours_Full.pdf.

*Monsaas, J. (1994). *Evaluation report–Final validation: Project EMERGE, Cript County Atlanta*, GA: Emory University.

*Morrison, G. M., Storino, M. H., Robertson, L. M., Weissglass, T., & Dondero, A. (2000). The protective function of after-school programming and parent education and support for students at risk for substance abuse. *Evaluation and Program Planning, 23*, 365–371.

National Research Council and Institute of Medicine. (2002). *Community programs to promote youth development*. Washington, DC: National Academy Press.

*Neufeld, J., Smith, M. G., Estes, H., & Hill, G. C. (1995). Rural after school child care: A demonstration project in a remote mining community. *Rural Special Education Quarterly, 14,* 12–16.

*Oyserman, D., Terry, K., & Bybee, D. (2002). A possible selves intervention to enhance school involvement. *Journal of Adolescence, 25*, 313–326.

Pechman, E. M., Russell, C. A., & Birmingham, J. (2008). *Out-of-school time (OST) observation instrument: Report of the validation study*. Washington, DC: Policy Studies Associates, Inc. Retrieved July 30, 2009 from www.policystudies.com.

*Philliber, S., Kaye, J., & Herrling, S. (2001). *The national evaluation of the Children's Aid Society Carrera model program to prevent teen pregnancy*. Accord, NY: Philliber Research Associates.

*Phillips, Ruby S. C. (1999). Intervention with siblings of children with developmental disabilities from economically disadvantaged families. *Families in Society, 80,* 569–577.

*Pierce, L. H., & Shields, N. (1998). The Be A Star community-based after-school program: Developing resiliency in high-risk preadolescent youth. *Journal of Community Psychology, 26,* 175–183.

*Prenovost, J. K. E. (2001). A first-year evaluation of after school learning programs in four urban middle schools in the Santa Ana Unified School District. *Dissertation Abstracts International*, 62 (03), 884A.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Riggs, N. R., & Greenberg, M. T. (2004). After-school youth development programs: A developmental-ecological model of current research. *Clinical Child and Family Review, 7*, 177–190.

*Ross, J. G., Saavadra, P. J., Shur, G. H., Winters, F., & Felner, R. D. (1992). The effectiveness of an after-school program for primary grade latchkey students on precursors of substance abuse. *Journal of Community Psychology, OSAP Special Issue,* 22–38.

Roth, J. L., Malone, L. M., & Brooks-Gunn, J. (2010). Does the amount of participation in afterschool programs relate to developmental outcomes: A review of the literature. *American Journal of Community Psychology*. doi:10.1007/s10464-010-9303-3.

*Rusche, S., Kemp, P., Krizmanich, J., Bowles, E., Moore, B., Craig Jr., H. E., et al. (1999). *Helping everyone reach out: Club Hero, final report*. Atlanta, GA: National Families in Action & Emstar Research.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52*, 471–499.

*Schinke, S. P., Orlandi, M. A., Botvin, G. J., Gilchrist, L. D., Trimble, J. E., & Locklear, V. S. (1988). Preventing substance abuse among American-Indian adolescents: A bicultural competence skills approach. *Journal of Counseling Psychology, 35,* 87–90.

*Schinke, S. P., Orlandi, M. A., & Cole, K. C. (1992). Boys & Girls Clubs in public housing developments: Prevention services for youth at risk. *Journal of Community Psychology, OSAP Special Issue*, 118–128.

Sheldon, J. Arbreton, A., Hopkins, L., & Grossman, J. B. (in press). Investing in success: Key strategies for building quality in after-school programs. *American Journal of Community Psychology*. doi:10.1007/s10464-010-9296-y.

Shernoff, D. J. (2010). Engagement in after-school programs as a predictor of social competence and academic performance. *American Journal of Community Psychology*. doi:10.1007/s10464-010-9314-0.

Simpkins, S., Little, P., & Weiss, H. (2004). *Understanding and measuring attendance in out-of-school time programs.* Cambridge, MA: Harvard Family Research Project. Available at www.gse.harvard.edu/hfrp/projects/afterschool/resources/issuebrief7.html.

*Smith, R. E., Smoll, F. L., & Curtis, B. (1979). Coach effectiveness training: A cognitive-behavioral approach to enhancing relationship skills in youth sport coaches. *Journal of Sport Psychology, 1*, 59–75.

*Smoll, F. L., Smith, R. E., Barnett, N. P., & Everett, J. J. (1993). Enhancement of children's self-esteem through social support training for youth sport coaches. *Journal of Applied Psychology, 78*, 602–610.

Southwest Educational Development Laboratory (2006). Time for Achievement: Afterschool and out-of-school time. *SEDL Letter, 18,* Number 1. Retrieved January 10, 2006 from http://www.sedl.org/pubs/sedl-letter/v18n01/SEDLLetter-v18n01.pdf.

*St. Pierre, T. L., & Kaltreider, D. L. (1992). Drug prevention in a community setting: A longitudinal study of the relative effectiveness of a three-year primary prevention program in Boys & Girls Club across the nation. *American Journal of Community Psychology, 20,* 673–706.

*St. Pierre, T. L., Mark, M. M., Kaltreider, D. L., & Aikin, K. J. (1997). Involving parents of high-risk youth in drug prevention: A three-year longitudinal study in Boys & Girls Clubs. *Journal of Early Adolescence, 17*, 21–50.

*St. Pierre, T. L., Mark, M. M., Kaltreider, D. L., & Campbell, B. (2001). Boys and Girls Clubs and school collaborations: A longitudinal study of a multicomponent substance abuse prevention program for high-risk elementary school children. *Journal of Community Psychology, 29*, 87–106.

*Tebes, J. K., Feinn, R., Vanderploeg, J. J. Chinman, M. J., Shepard, J. Brabham, T., et al. (2007). Impact of a positive youth development program in urban after-school settings on the prevention of adolescent substance use. *Journal of Adolescent Health, 41,* 239–247.

Tobler, N. S., Roona, M. R., Ochshorn, P., Marshall, D. G., Streke, A. V., & Stackpole, K. M. (2000). School-based adolescent drug prevention programs: 1998 meta-analysis. *The Journal of Primary Prevention, 20*, 275–336.

*Tucker, C. M., & Herman, K. C. (2002). Using culturally sensitive theories and research to meet the academic needs of low-income African American children. *American Psychologist, 57,* 762-773.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret effect sizes. *Journal of Counseling Psychology, 51*, 473–481.

*Vandell, D. L., Reisner, E. R., Brown, B. B., Dadisman, K., Pierce, K. M., & Lee, D. et al. (2005). *The study of promising after-school programs: Examination of intermediate outcomes in year 2*. Retrieved June 16, 2006, from http://www.wcer.wisc.edu/childcare/statements.html.

*Vandell, D. L., Reisner, E. R., Brown, B. B., Dadisman, K., Pierce, K. M., & Lee, D. (2004). *The study of promising after-school programs: Descriptive report of the promising programs*. University of Wisconsin, Madison: Wisconsin Center for Education Research. Retrieved June 16, 2006, from http://www.wcer.wisc.edu/childcare/statements.html.

*Vincent, V., & Guinn, R. (2001). Effectiveness of a Colonia educational intervention. *Hispanic Journal of Behavioral Sciences, 23*, 229–238.

*Weisman, S. A., Soule, D. A., & Womer, S. C. (2001). *Maryland after school community grant program report on the 1999–2000 school year evaluation of the phase 1 after-school programs*. University of Maryland, College Park.

*Weisman, S. A., Womer, S. C., Kellstrom, M., Bryner, S., Kahler, A., Slocum, L. A., et al. (2003). *Maryland after school grant program part 1: Report on the 2001–2002 school year evaluation of the phase 3 after school programs*. University of Maryland, College Park.

Weissberg, R. P., & Greenberg, M. T. (1998). School and community competence-enhancement and prevention programs. In I. E. Siegel & K. A. Renninger (Eds.), *Handbook of child psychology. Vol 4: Child psychology in practice* (5th ed., pp. 877–954). New York: Wiley.

Wilson, D. B., Gottfredson, D. C., & Najaka, S. S. (2001). School-based prevention of problem behaviors: A meta-analysis. *Journal of Quantitative Criminology, 17*, 247–272.

Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology, 71*, 136–149.

Yohalem, N., Wilson-Ahlstrom, A. Fischer, S., & Shinn, M. (2007). *Measuring youth program quality: A guide to assessment tools.* Retrieved April 20, 2007 from www.forumfyi.org/Files/Measuring_youth_ program_quality.pdf.

*Zief, S. G. (2005). *A mixed methods study of the impacts and processes of an after-school program for urban elementary youth*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.

Zins, J. E., Weissberg, R. P., Wang, M. C., & Walberg, H. J. (Eds.). (2004). *Building academic success on social and emotional learning: What does the research say?*. New York: Teachers College Press.



View publication stats