Early Social-Emotional Functioning and Public Health: The Relationship Between Kindergarten Social Competence and Future Wellness

Damon E. Jones, PhD, Mark Greenberg, PhD, and Max Crowley, PhD

Understanding what early characteristics predict future outcomes could be of great value in helping children develop into healthy adults. In recent years, much research has been directed toward understanding noncognitive traits in children that may increase the likelihood of healthy personal development and eventual adult well-being. For predicting future success in the workplace, levels of cognitive ability measured through IQ or test scores alone are less predictive than measures of educational attainment, which require not just cognitive ability but also noncognitive characteristics such as self-discipline, academic motivation, and interpersonal skills. Future likelihood of committing crimes is greatly influenced by noncognitive processes in development, such as externalizing behavior, social empathy, and effectively regulating emotions. A recent study found that noncognitive ability in the form of self-control in childhood was predictive of adult outcomes ranging from physical health to crime to substance abuse. The value of noncognitive skills has also been determined through evaluation of interventions such as the landmark Perry Preschool program, in which improvements in noncognitive skills related to behavior and academic motivation were found to be central to long-term effects on crime and employment.

Inadequate levels of social and emotional functioning are increasingly recognized as central to many public health problems (e.g., substance abuse, obesity, violence). Just as researchers study how academic achievement in a population can lift groups out of poverty, public health scientists are now studying how these noncognitive factors affect health and wellness across domains.

Classification of characteristics into complementary cognitive and noncognitive categories is a convenient way to characterize competencies in human development. Cognitive skills involve achievement-oriented tasks, such as problem solving, and academic abilities, which are measured by achievement tests; the noncognitive category covers everything else, such as behavioral characteristics, emotion regulation, attention, self-regulation, and social skills. Designation of cognitive versus noncognitive skills oversimplifies the complexity of skills and the role of cognition. Cognitive skills are involved not only in intelligence and achievement, but also in attention, emotion regulation, attitudes, motivation, and the conduct of social relationships (e.g., Farrington et al. provide an overview of noncognitive traits in educational research).

Noncognitive skills interact with cognitive skills to enable success in school and the workplace. This is most easily seen in an educational setting. Achievement is driven by intellectual ability as well as by the self-regulation, positive attitudes, motivation, and conscientiousness that are required to complete educational milestones. Substantial differences in noncognitive skills have been found between those who graduate from high school on time and those who complete a general equivalency diploma, as reflected in subsequent adult and economic outcomes.

Interpersonal skills are also important for children navigating the social setting, and positive interactions with adults are essential for success in school. Success in school involves both social-emotional and cognitive skills, because social interactions, attention, and self-control affect readiness for learning.

An additional feature of noncognitive competencies is that they may be more malleable than cognitive skills and thus may be appropriate targets for prevention or intervention efforts. Of course, the degree to which this is true depends on the specific skill and on...
multiple factors associated with children’s characteristics and environment. Regardless, a challenge lies in effectively assessing children’s competencies at an early enough age that intervention or prevention efforts might be introduced. Although an assessment at any 1 point may be inadequate for summarizing an individual’s overall noncognitive competencies, it is useful to know what early competencies predict future success and avoidance of problems. This is especially relevant in light of studies showing the value of enhancing the social-behavioral and learning environment of young children, to foster positive child development as well as to alter adult health and labor market outcomes.

A key characteristic of noncognitive ability in young children is social competence. Social competence encompasses both the ability to complete tasks and manage responsibilities and effective skills for handling social and emotional experiences. Children’s social competence can be assessed by their kindergarten teachers, who observe many instances in which children need to manage relations with peers and adults. The school setting provides the opportunity to observe children’s abilities to interact interpersonally as they cooperate with others to complete daily tasks and resolve conflicts. Such skills are important for successful progression in early grades.

We investigated how well key late adolescent and early adult outcomes were predicted by teacher ratings of children’s social competence (1 indicator of early noncognitive ability) measured many years previously in kindergarten, to foster positive child development as well as to alter adult health and labor market outcomes.

For predictors we focused on the earliest age for which data were available: measures obtained when children were in kindergarten. Throughout the analytic process we found it useful to consider whether other important background variables predicted future outcomes. However, our primary objective was to determine how well an inexpensive, easily obtained snapshot of social competence at formal entrance to school predicted important outcomes, after adjustment for other expected influences on development, such as family circumstances, gender, academic ability, and behavior. If such a measure can identify early noncognitive deficiencies, this could provide important information for determining potential targets for early intervention.

METHODS

We used data from the longitudinal, non-intervention subsample of the Fast Track Project, an intervention program designed to reduce aggression in children identified as at high risk for long-term behavioral problems and conduct disorders. The Fast Track study design comprised an intervention group and a matched control group sample of high-risk children as well as a non–high-risk (normative) subsample of students attending control schools. We focused on the high-risk control students and the normative sample—those individuals who did not receive any Fast Track prevention services. The total sample size was 753 (high-risk control group, n = 367; non–high-risk, normative group, n = 386).

Participants were recruited from the 4 study sites (3 urban, 1 rural): Durham, North Carolina; Nashville, Tennessee; Seattle, Washington; and central Pennsylvania. Further information on the Fast Track Project sample recruitment process is available in study publications. In the total sample, 58% were boys, about 50% were White, 46% were African American, and 4% had other racial/ethnic backgrounds. The study oversampled higher-risk students, and we employed sampling probability weights in all analyses. More information on the design is provided in Appendix A, which describes the screening and recruitment process (available as a supplement to the online version of this article at http://www.ajph.org).

The project first collected data when children were attending kindergarten; initial data collection for the first cohort took place in 1991. Final follow-up data were collected 19 years later, when participants were aged approximately 25 years. Participation from the original sample was high, and we found no differential response in analyses considering a range of baseline variables. More detail on this assessment and the follow-up sample are provided in a recent study of long-term intervention effects.

Our outcome measures concerned education, employment, public assistance, crime, mental health, and substance use. The project measured all outcomes through late adolescence or early adulthood. We included relevant background variables in the models to control for characteristics of the children at kindergarten age and their families. Most importantly, we selected control variables that would better enable identification of unique prediction attributable to early social skills. Thus, models included variables representing family demographics (gender, race, number of parents in the home, socioeconomic status), early childhood aggression (both in school and at home), early academic ability, and other contextual factors. We did not include the indicator for gender in models of justice system outcomes because of the very low rate of criminal offenses among female participants. We did not include region as a covariate in models. Initially we included 3 dummy variables to represent project site, but we removed this covariate when initial tests indicated little difference between regions on the study outcome variables.

Table 1 provides the outcomes and control variables for all analytic models, with information on the scales used and the data sources. Appendix A (available as an online supplement) provides more details on measurement sources and scale reliabilities for all variables used in analyses.

To represent social competence in kindergarten, we chose the Prosocial–Communication Skills subscale of the Social Competence Scale. The score combined 9 items that teachers rated on a 5-point Likert scale, assessing how the child interacted socially with others. Examples
<table>
<thead>
<tr>
<th>Variable</th>
<th>Survey</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model outcomes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education/employment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school graduation on time</td>
<td>National Longitudinal Surveys</td>
<td>Self-report</td>
</tr>
<tr>
<td>College graduation</td>
<td>Self-report</td>
<td></td>
</tr>
<tr>
<td>Currently employed full-time</td>
<td>Self-report</td>
<td></td>
</tr>
<tr>
<td>Stable employment</td>
<td>Self-report</td>
<td></td>
</tr>
<tr>
<td>Years of special education services, no.</td>
<td>School Archival Records Survey (grades 1-12)</td>
<td>School records</td>
</tr>
<tr>
<td>Years of repeated grades, no.</td>
<td>School records</td>
<td></td>
</tr>
<tr>
<td><strong>Public assistance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On waiting list for public housing</td>
<td>Neighborhoods and Government Programs</td>
<td>Self-report</td>
</tr>
<tr>
<td>Receiving public assistance</td>
<td>Self-report</td>
<td></td>
</tr>
<tr>
<td>Receiving unemployment compensation</td>
<td>Self-report</td>
<td></td>
</tr>
<tr>
<td><strong>Crime</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrests for severe offense, no.</td>
<td>Juvenile and adult court data</td>
<td>Court records</td>
</tr>
<tr>
<td>Ever arrested</td>
<td>Service Assessment for Children and Adolescents</td>
<td>Self-report</td>
</tr>
<tr>
<td>Ever arrested</td>
<td>Self-report</td>
<td></td>
</tr>
<tr>
<td>Ever made court appearance</td>
<td>Self-report</td>
<td></td>
</tr>
<tr>
<td>Ever made court appearance</td>
<td>Self-report</td>
<td></td>
</tr>
<tr>
<td>Ever had police contact</td>
<td>Self-report</td>
<td></td>
</tr>
<tr>
<td>Ever stayed in detention facility</td>
<td>Combined outcomes from self-report and criminal records</td>
<td>Self-report, court records</td>
</tr>
<tr>
<td><strong>Substance abuse</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol dependence</td>
<td>Self-Reported Substance Use and Dependence</td>
<td>Self-report</td>
</tr>
<tr>
<td>Drug dependence</td>
<td>Self-report</td>
<td></td>
</tr>
<tr>
<td>Smoked regularly in past month</td>
<td>Tobacco, Alcohol and Drugs survey</td>
<td>Self-report</td>
</tr>
<tr>
<td>Days of binge drinking in past month, no.</td>
<td>Self-report</td>
<td></td>
</tr>
<tr>
<td>Days of marijuana use in past month, no.</td>
<td>Self-report</td>
<td></td>
</tr>
<tr>
<td><strong>Mental health</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Externalizing problems</td>
<td>Young Adult Self-Report</td>
<td>Self-report</td>
</tr>
<tr>
<td>Internalizing problems</td>
<td>Social Health Profile</td>
<td>Primary caregiver</td>
</tr>
<tr>
<td>Years on medications, no.</td>
<td>Primary caregiver</td>
<td></td>
</tr>
<tr>
<td><strong>Model predictors (for child at kindergarten age)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (female)</td>
<td>Family information form</td>
<td>Primary caregiver</td>
</tr>
<tr>
<td>Race (African American)</td>
<td>Primary caregiver</td>
<td></td>
</tr>
<tr>
<td>Family socioeconomic status (Hollingshead code)</td>
<td>Primary caregiver</td>
<td></td>
</tr>
<tr>
<td>Mother an adolescent at child’s birth</td>
<td>Primary caregiver</td>
<td></td>
</tr>
<tr>
<td>Neighborhood total score</td>
<td>Neighborhood Questionnaire</td>
<td>Primary caregiver</td>
</tr>
<tr>
<td>Life stresses total score</td>
<td>Life Changes Survey</td>
<td>Primary caregiver</td>
</tr>
<tr>
<td>Letter-word identification score</td>
<td>Woodcock-Johnson Psycho-Educational Battery</td>
<td>Administered survey</td>
</tr>
<tr>
<td>Authority acceptance</td>
<td>Teacher Observation of Child Adaptation-Rev</td>
<td>Teacher</td>
</tr>
<tr>
<td>Externalizing score</td>
<td>Child Behavior Checklist</td>
<td>Primary caregiver</td>
</tr>
<tr>
<td>Prosocial-communication skills</td>
<td>Social Competence Scale</td>
<td>Teacher</td>
</tr>
</tbody>
</table>

*a*Measured at age 25 years.  
*b*Through high school.  
*c*Measured after high school (aged 19-20 years).  
*d*For emotional or behavioral issues.
of these items include “cooperates with peers without prompting,” “is helpful to others,” “very good at understanding feelings,” and “resolves problems on own.” Internal reliability coefficients were very high (α = 0.97), and univariate assessment demonstrated good distributional characteristics (unweighted mean = 1.90; SD = 0.97). The subscale was highly associated with other subscales in the measure, such as the Emotion Regulation subscale (r = 0.90).

A natural question in this type of research is whether associations may differ because of differing background variables. Although we did not formally investigate potential moderation of associations, we explored whether race or gender moderated links within domains. We executed a representative number of models from each domain with an interaction term entered for the cross between the potential moderator and prosocial skills. In this preliminary investigation, we found no patterns of moderation exerted by race or gender on any outcome domains. We therefore did not conduct extensive tests of moderation (to keep the number of statistical tests for overall models manageable). Follow-up research could include a more specific focus on the potential differences in linkages within a given outcome domain across key demographic distinctions.

We used separate regression models for each study outcome. We regressed dependent variables on our control variables as well as on the social competence score. We ran logistic regressions for all dichotomous outcomes and count-based regressions for the measures of amounts. The latter involved Poisson regression unless outcomes were overdistributed, in which case we used a negative binomial modeling specification. We used a zero-inflated Poisson model for 1 count outcome (number of arrests for severe crimes by age 25 years). We conducted analyses with M-Plus software with full-information maximum likelihood estimation techniques, which provided results representing the full sample (n = 753) at kindergarten (integrating over the missing cases). We used Monte Carlo integration techniques for parameter estimates, because of the categorical nature of the outcomes. We also specified robust standard error estimation for all models.

Rates of missing data varied by outcome (Table 2). Attrition was lower for outcomes obtained prior to the end of high school. Missing data rates also were lower for outcomes obtained through public criminal records at early adulthood. Accommodation of missing data through full-information maximum likelihood procedures assumes that missing data are conditionally missing at random, with all measured covariates in the analytic model considered.

**RESULTS**

Table 2 provides the means and standard deviations for predictors in all analytic models and for the separate adolescent and adult outcomes that we examined. Results from regression models are presented in Table 3 for the estimate on prosocial skills. Odds ratios (ORs) are provided for results from logistic regression models; incidence rate ratios (IRRs) are provided for results from count-based regression models. We considered results significant at P < 0.05. Appendix B (available as a supplement to the online version of this article at http://www.ajph.org) shows statistical significance results for all model covariates and details on joint prediction among all variables; estimates are indicated in terms of direction of association.

Our analyses included 4 education and employment outcomes representing attainment through age 25 years. Kindergarten prosocial skills were significantly and uniquely predictive of all 4 outcomes: whether participants graduated from high school on time (OR = 1.54; 95% confidence interval [CI] = 1.09, 2.19; P < 0.05; Table 3), completed a college degree (OR = 2.00; 95% CI = 1.07, 3.75; P < 0.05), obtained stable employment in young adulthood (OR = 1.66; 95% CI = 1.13, 2.43; P < 0.01), and were employed full time in young adulthood (OR = 1.46; 95% CI = 1.02, 2.08; P < 0.05). For the 2 outcomes spanning school ages, we observed a negative association for number of years of special education services (IRR = 0.54; 95% CI = 0.44, 0.67; P < 0.001) and number of years of repeated grades through high school (IRR = 0.79; 95% CI = 0.65, 0.97; P < 0.05). Two of the 3 outcomes representing public assistance in young adulthood were significantly linked to early social competence. Early prosocial skills were negatively related to the likelihood of living in or being on a waiting list for public housing (OR = 0.55; 95% CI = 0.36, 0.85; P < 0.01; Table 3) and of receiving public assistance (OR = 0.63; 95% CI = 0.43, 0.91; P < 0.05). We found no significant association for receiving unemployment compensation in young adulthood.

Results for justice system outcomes demonstrated consistent patterns across different ages and variables. Early prosocial skills were significantly inversely predictive of any involvement with police before adulthood (OR = 0.65; 95% CI = 0.45, 0.94; P < 0.05) and ever being in a detention facility (OR = 0.61; 95% CI = 0.40, 0.94; P < 0.05). Although juveniles’ self-report of whether they had been arrested and or had appeared in court followed the same pattern, the estimates were not statistically significant at conventional levels. In young adulthood, early social competence was significantly and uniquely linked to being arrested (OR = 0.60; 95% CI = 0.44, 0.90; P < 0.05) and appearing in court (OR = 0.63; 95% CI = 0.43, 0.91; P < 0.05). Finally, early social competence significantly predicted the number of arrests for a severe offense by age 25 years (IRR = 0.68; 95% CI = 0.49, 0.94; P < 0.05), as determined through public records.

Although early social competence was not associated with alcohol and drug dependence diagnoses in early adulthood, our models showed that it correlated with substance abuse behavior. We found statistically significant associations in separate models of the number of days of binge drinking in the past month (IRR = 0.66; 95% CI = 0.44, 0.97; P < 0.05) and the number of days marijuana was used (IRR = 0.55; 95% CI = 0.35, 0.87; P < 0.01). An association with regular tobacco use was not significant.

Results were mixed on associations between early prosocial skills and future mental health outcomes, although patterns were consistent with findings in other domains. Links between kindergarten prosocial skills and future internalizing and externalizing problems were nonsignificant at conventional levels. Finally, early prosocial skills significantly predicted number of years on medication for emotional or behavioral issues through high school (OR = 0.54; 95% CI = 0.40, 0.75; P < 0.001).
DISCUSSION

We examined whether early childhood social competence predicted outcomes measured up to 2 decades later. We evaluated outcomes that broadly represented personal well-being, covering domains of education, employment, crime, substance use, and mental health. Such outcomes are markers of personal success or avoidance of problems. These outcomes are also economically relevant to both individual and public resources. Overall, results indicated statistically significant and unique associations between teacher-assessed prosocial skills and outcomes in all domains examined.

We used a rich database that combined a long time frame of data collection with coverage of various domains of human development and adult outcomes. Such data provided the unique opportunity to investigate the importance of early social-emotional characteristics. An additional strength of these data was that they involved multiple sources of information: teachers, parents, self-reports, and public records.

Our results support previous research that examined long-term prediction from noncognitive skills, most notably Moffitt et al., who found that self-control across early childhood was a significant predictor of outcomes in multiple domains of early adult functioning.4 Other important research has shown that noncognitive skills are not as reliable predictors for some outcomes (e.g., achievement), as other, more strictly cognitive characteristics, such as academic achievement at school entry.10,35 Results across studies likely differ because of variation in predictors used, quality of measurement of study constructs, outcome domains, age at baseline and follow-up, and other characteristics of the population studied.

Our results demonstrate the predictive power of teacher-measured prosocial skills independent of child, family, and contextual factors that typically predict adult outcomes, because we controlled for socioeconomic status, family risk status, neighborhood quality, and children’s characteristics (notably behavioral traits and early academic ability). Our results confirm that these control variables are indeed predictive of some adult outcomes but

### TABLE 2—Model Predictors and Adolescent and Young Adult Outcomes Associated With Social-Emotional Functioning in Kindergarten: Fast Track Project, United States, 1991–2010

<table>
<thead>
<tr>
<th>Variable</th>
<th>No.</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model predictors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (female)</td>
<td>753</td>
<td>0.42 (0.49)</td>
</tr>
<tr>
<td>Race (African American)</td>
<td>753</td>
<td>0.46 (0.50)</td>
</tr>
<tr>
<td>Family socioeconomic status (Hollingshead code)</td>
<td>753</td>
<td>25.05 (12.90)</td>
</tr>
<tr>
<td>Mother an adolescent at child’s birth</td>
<td>636</td>
<td>0.16 (0.37)</td>
</tr>
<tr>
<td>Neighborhood total score</td>
<td>752</td>
<td>0.03 (0.61)</td>
</tr>
<tr>
<td>Life stresses total score</td>
<td>745</td>
<td>1.51 (0.75)</td>
</tr>
<tr>
<td>Woodcock-Johnson letter-word identification score</td>
<td>752</td>
<td>12.83 (4.22)</td>
</tr>
<tr>
<td>Authority acceptance (teacher-rated behavior)</td>
<td>749</td>
<td>57.34 (11.57)</td>
</tr>
<tr>
<td>Child Behavior Checklist externalizing score (parent-rated behavior)</td>
<td>746</td>
<td>57.57 (10.20)</td>
</tr>
<tr>
<td>Prosocial–communication skills</td>
<td>686</td>
<td>1.90 (0.97)</td>
</tr>
<tr>
<td><strong>Model outcomes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education/employment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school graduation on time</td>
<td>620</td>
<td>0.63 (0.48)</td>
</tr>
<tr>
<td>College graduation</td>
<td>620</td>
<td>0.11 (0.32)</td>
</tr>
<tr>
<td>Currently employed full-time</td>
<td>621</td>
<td>0.49 (0.50)</td>
</tr>
<tr>
<td>Stable employment</td>
<td>611</td>
<td>0.32 (0.47)</td>
</tr>
<tr>
<td>Years of special education services, no.</td>
<td>736</td>
<td>2.19 (3.56)</td>
</tr>
<tr>
<td>Years of repeated grades, no.</td>
<td>751</td>
<td>0.66 (0.90)</td>
</tr>
<tr>
<td>Public assistance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>On waiting list for public housing</td>
<td>615</td>
<td>0.16 (0.37)</td>
</tr>
<tr>
<td>Receiving public assistance</td>
<td>603</td>
<td>0.34 (0.47)</td>
</tr>
<tr>
<td>Receiving unemployment compensation</td>
<td>603</td>
<td>0.18 (0.38)</td>
</tr>
<tr>
<td>Crime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrests for severe offense, no.</td>
<td>753</td>
<td>0.12 (0.33)</td>
</tr>
<tr>
<td>Ever arrested</td>
<td>516</td>
<td>0.34 (0.47)</td>
</tr>
<tr>
<td>Ever arrested</td>
<td>525</td>
<td>0.26 (0.44)</td>
</tr>
<tr>
<td>Ever made court appearance</td>
<td>519</td>
<td>0.35 (0.48)</td>
</tr>
<tr>
<td>Ever made court appearance</td>
<td>534</td>
<td>0.33 (0.47)</td>
</tr>
<tr>
<td>Ever had police contact</td>
<td>562</td>
<td>0.60 (0.49)</td>
</tr>
<tr>
<td>Ever stayed in detention facility</td>
<td>526</td>
<td>0.22 (0.42)</td>
</tr>
<tr>
<td>Substance abuse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol dependence</td>
<td>556</td>
<td>0.26 (0.44)</td>
</tr>
<tr>
<td>Drug dependence</td>
<td>550</td>
<td>0.10 (0.30)</td>
</tr>
<tr>
<td>Smoked regularly in past month</td>
<td>575</td>
<td>0.38 (0.49)</td>
</tr>
<tr>
<td>Days of binge drinking in past month, no.</td>
<td>602</td>
<td>1.69 (4.65)</td>
</tr>
<tr>
<td>Days of marijuana use in past month, no.</td>
<td>607</td>
<td>3.60 (8.94)</td>
</tr>
<tr>
<td>Mental health</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Externalizing problems</td>
<td>620</td>
<td>0.21 (0.41)</td>
</tr>
<tr>
<td>Internalizing problems</td>
<td>620</td>
<td>0.29 (0.46)</td>
</tr>
<tr>
<td>Years on medications, no.</td>
<td>720</td>
<td>0.83 (2.14)</td>
</tr>
</tbody>
</table>

Note. Participants were recruited from 4 study sites (3 urban, 1 rural): Durham, NC; Nashville, TN; Seattle, WA; and central Pennsylvania.

aAt age 25 years.
bThrough high school.
cMeasured after high school (aged 19–20 years).
dFor emotional or behavioral issues.
that additional, unique variance can be attributed to social competence at school entrance. In many cases, social competence was a stronger predictor (according to statistical $P$ values) than factors seemingly more directly aligned with the outcome. This was most striking in our comparison of associations of kindergarten teacher–rated aggression and social competence with later crime outcomes: the measure of prosocial skills was a consistent predictor of future crime outcomes, but the level of aggression observed by the same teacher was not usually significantly predictive after adjustment for other factors (including a separate measure of aggression from the primary caretaker). A partial explanation may be that aggression is a less stable characteristic among kindergarteners than is the broader domain of positive social relations. Furthermore, although a relatively small percentage of children show early aggressive behavior and, thus, skew the distribution, social competence is more normally distributed and therefore may be a better predictor across the spectrum. The 2 measures shared the same rater and were moderately correlated (roughly 0.50 in this sample), as would be expected.

**Limitations**

Focusing on a single measurement at an early age is somewhat risky because characteristics of social competence as recognized by teachers may manifest in different ways in later years. We could not determine causal associations, but our findings suggest the potential for such a measure to be used in screening for intervention at an early stage of development. Noncognitive factors such as conscientiousness, self-regulation, motivation, academic ability, and other attitudes and behaviors in later childhood years may be more important markers of long-term outcomes, but they have not yet been fully developed and thus have not been efficiently assessed in children at 5 years of age.

Our measure of social competence was a continuous composite from teacher observation that combined multiple social-behavioral scenarios for the child. This measure, although subject to measurement error, likely represents children’s social competence relatively well, because the teacher has been a daily observer in the classroom setting. For the kindergarten data, we were not able to clearly distinguish between social competence and self-regulation, because the 2 scales were so highly correlated (and thus were not included in the same multiple regression). Self-regulation is likely reflected in socially competent behavior but is multidimensional and may be assessed independently through tests of executive function as children mature and take on more responsibility to progress through school.

Our measure of social competence was continuous, raising the issue of whether there may be certain cutoffs (e.g., very low competence) where this characteristic might be especially predictive of later outcomes. In addition, with the available data, we were not able to assess the validity of the measure for prosocial skills. We focused on what was measured at

---

**TABLE 3—Logistic Regression and Negative Binomial Regression Results for Associations of Social-Emotional Functioning in Kindergarten With Adolescent and Young Adult Outcomes: Fast Track Project, United States, 1991–2010**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>OR (95% CI)</th>
<th>IRR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education/employment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school graduation on time$^a$</td>
<td>1.54* (1.09, 2.19)</td>
<td></td>
</tr>
<tr>
<td>College graduation$^a$</td>
<td>2.00* (1.07, 3.75)</td>
<td></td>
</tr>
<tr>
<td>Currently employed full-time$^a$</td>
<td>1.46* (1.02, 2.08)</td>
<td></td>
</tr>
<tr>
<td>Stable employment$^a$</td>
<td>1.66** (1.13, 2.43)</td>
<td></td>
</tr>
<tr>
<td>Years of special education services$^b$, no.</td>
<td>0.54*** (0.44, 0.67)</td>
<td></td>
</tr>
<tr>
<td>Years of repeated grades, no.</td>
<td>0.79* (0.65, 0.97)</td>
<td></td>
</tr>
<tr>
<td>Public assistance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Living in/on waiting list for public housing$^c$</td>
<td>0.55** (0.36, 0.85)</td>
<td></td>
</tr>
<tr>
<td>Receiving public assistance$^d$</td>
<td>0.63* (0.43, 0.91)</td>
<td></td>
</tr>
<tr>
<td>Receiving unemployment compensation$^e$</td>
<td>0.89 (0.55, 1.45)</td>
<td></td>
</tr>
<tr>
<td>Crime</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrests for severe offense$^f$, no.</td>
<td>0.68* (0.49, 0.94)</td>
<td></td>
</tr>
<tr>
<td>Ever arrested$^g$</td>
<td>0.67 (0.44, 1.02)</td>
<td></td>
</tr>
<tr>
<td>Ever arrested$^h$</td>
<td>0.60* (0.40, 0.90)</td>
<td></td>
</tr>
<tr>
<td>Ever made court appearance$^i$</td>
<td>0.70 (0.47, 1.03)</td>
<td></td>
</tr>
<tr>
<td>Ever made court appearance$^j$</td>
<td>0.63* (0.43, 0.91)</td>
<td></td>
</tr>
<tr>
<td>Ever had police contact$^k$</td>
<td>0.65* (0.45, 0.94)</td>
<td></td>
</tr>
<tr>
<td>Ever stayed in detention facility$^{b,c}$</td>
<td>0.61* (0.40, 0.94)</td>
<td></td>
</tr>
<tr>
<td>Substance abuse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol dependence$^m$</td>
<td>0.89 (0.59, 1.35)</td>
<td></td>
</tr>
<tr>
<td>Drug dependence$^n$</td>
<td>0.86 (0.45, 1.65)</td>
<td></td>
</tr>
<tr>
<td>Smoked regularly in past month$^o$</td>
<td>0.71 (0.48, 1.04)</td>
<td></td>
</tr>
<tr>
<td>Days of binge drinking in past month, no.</td>
<td>0.66* (0.44, 0.97)</td>
<td></td>
</tr>
<tr>
<td>Days of marijuana use in past month, no.</td>
<td>0.55** (0.35, 0.84)</td>
<td></td>
</tr>
<tr>
<td>Mental health</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Externalizing problems$^p$</td>
<td>0.61 (0.36, 1.02)</td>
<td></td>
</tr>
<tr>
<td>Internalizing problems$^q$</td>
<td>0.70 (0.48, 1.03)</td>
<td></td>
</tr>
<tr>
<td>Years on medications$^{b,d}$, no.</td>
<td>0.54*** (0.40, 0.75)</td>
<td></td>
</tr>
</tbody>
</table>

Note: CI = confidence interval; IRR = incidence rate ratio; OR = odds ratio. Participants were recruited from 4 study sites (3 urban, 1 rural): Durham, NC; Nashville, TN; Seattle, WA; and central Pennsylvania. Control variables were gender, race (African American), family socioeconomic status, neighborhood quality, family life stressors, whether mother was an adolescent, early academic skill, teacher-rated aggression, and parent-rated aggression.

$^a$At age 25 years.

$^b$Through high school.

$^c$Measured after high school (aged 19–20 years).

$^d$For emotional or behavioral issues.

$^e$For emotional or behavioral issues.

$^f$P < .05; **P < .01; ***P < .001.
school entry and likely fell well short of completely understanding noncognitive skills and what it might entail throughout development.

**Conclusions**

Our goal was to examine what can be assessed at school entrance when plans for addressing problems or enhancing skills may best be initiated. Our results suggest that perceived early social competence at least serves as a marker for important long-term outcomes and at most is instrumental in influencing other developmental factors that collectively affect the life course. Evaluating such characteristics in children could be important in planning interventions and curricula to improve these social competencies. Although “softer” skills can be more malleable and, thus, possibly better candidates for intervention, they are also less likely to be captured in a single measurement at a single time than are variables such as IQ. Certainly, interventionists are challenged not only by what specific skills to focus on, but also by what ages to assess, how to consider the likely interactions with other traits (including cognitive skills), the role of contextual factors, and how best to measure (what sources, whether to combine measures, etc.).

The growing body of literature that demonstrates the importance of noncognitive skills in development should motivate policymakers and program developers to target efforts to improve these skills in young children. Much evidence has shown how effective intervention in preschool and the early elementary years can improve childhood noncognitive skills in a lasting way. Enhancing these skills can have an impact in multiple areas and therefore has potential for positively affecting individuals as well as community public health substantially.

Our study demonstrates the unique predictive nature of early social competence on important outcomes in late adolescence and early adulthood. Our results showed that teacher-rated prosocial skills in kindergarten were a consistently significant predictor across all outcome domains studied; thus, a measure such as this may be a good candidate for assessing whether children are at risk for deficits in noncognitive skills at school entry. We look forward to further research on the importance of social-emotional competencies in early development, especially among individuals more at risk for problems or less prepared to succeed in school or (eventually) the labor force. Such research ideally will advance understanding of the appropriate constructs and measures to focus on, with consideration of the age and context of the individual.

**About the Authors**

Damon E. Jones and Mark Greenberg are with the Bennett Pierce Prevention Research Center, Pennsylvania State University. University Park. Max Crowley is with the Center for Child and Family Policy, Duke University, Durham, NC.

Correspondence should be sent to Damon Jones. 310 Behavioral Health Building, The Pennsylvania State University. University Park, PA 16802 (e-mail: dej10@psu.edu). Reprints can be ordered at http://www.ajph.org by clicking the “Reprints” link.

This article was accepted February 10, 2015.

**Contributors**

D. E. Jones analyzed the data and was primary writer of the article. M. Greenberg helped plan data analyses and write the article. M. Crowley helped with analytic strategy and writing.

**Acknowledgments**

This study was funded by the Robert Wood Johnson Foundation (grant 70595, Damon E. Jones, principal investigator).

We are grateful to the Conduct Problems Prevention Research Group, Karen Bierman, PhD, John Coie, PhD, Ken Dodge, PhD, John Lochman, PhD, Robert McMahon, PhD, and Ellen Puderhughes, PhD, for providing the data. We also appreciate their feedback on an initial draft of the article.

**Human Participant Protection**

The Pennsylvania State University Institutional Review Board determined that no protocol approval was required because the study used secondary, de-identified data.

**References**


Effects of Social Development Intervention in Childhood 15 Years Later
J. David Hawkins, PhD; Rick Kosterman, PhD; Richard F. Catalano, PhD; Karl G. Hill, PhD; Robert D. Abbott, PhD

Abstract

Objective To examine the long-term effects of a universal intervention in elementary schools in promoting positive functioning in school, work, and community, and preventing mental health problems, risky sexual behavior, substance misuse, and crime at ages 24 and 27 years.

Design Nonrandomized controlled trial.

Setting Fifteen public elementary schools serving diverse neighborhoods including high-crime neighborhoods in Seattle, Washington.

Participants Sex-balanced and multiracial/multiethnic sample of 598 participants at ages 24 and 27 years (93% of the original sample in these conditions).

Interventions Teacher training in classroom instruction and management, child social and emotional skill development, and parent workshops.

Main Outcome Measures Self-reports of functioning in school, work, and community and of mental health, sexual behavior, substance use, and crime, and court records.

Results A significant multivariate intervention effect across all 16 primary outcome indices was found. Specific effects included significantly better educational and economic attainment, mental health, and sexual health by age 27 years (all \( P < .05 \)). Hypothesized effects on substance use and crime were not found at ages 24 or 27 years.

Conclusions A universal intervention for urban elementary schoolchildren, which focused on classroom management and instruction, children's social competence, and parenting practices, positively affected mental health, sexual health, and educational and economic achievement 15 years after the intervention ended.

Poverty, unemployment, and neighborhood disorganization are persistent problems in US cities.\(^1\)\(^2\) Crime, drug use, teen pregnancy, mental health problems, and high rates of school dropout plague many urban children and families.\(^3\)\(^4\) Public schools, available to all children in the United States beginning at age 5 or 6 years, are a potentially powerful setting for preventive intervention. We examined the effects of a 3-component preventive intervention provided in public schools during the elementary grades on outcomes at ages 24 and 27 years, 15 years after the intervention ended. The objective of the intervention was to improve the skills of teachers, parents, and children to increase positive functioning in school and decrease problems related to mental health, risky sexual behavior, substance use, and criminal behavior.

The mid-20s are important years for the adoption of adult roles. Engagement in education or occupational roles is an important predictor of future adult functioning.\(^5\)\(^6\) Civic engagement is also likely to increase during this period.\(^7\)\(^8\) However, the mid-20s are also years of relatively high vulnerability to mental health problems,\(^9\)\(^10\) sexual risk-taking,\(^11\)\(^12\) and continued risk of substance use and crime.\(^13\)\(^14\)

Little is known about the long-term effects of universal intervention in public elementary schools on these outcomes. Kellam et al.\(^15\)\(^16\) observed students from schools serving predominantly African American children from poor to lower middle-class families who had been exposed to a classroom-based behavior management program in the first and second grades. By ages 19 to 21, male subjects, particularly those who had demonstrated more aggressive or disruptive behavior in the first grade, reported significantly reduced rates of regular cigarette smoking, fewer drug and alcohol abuse or dependence disorders, and less antisocial personality disorder.\(^17\)\(^18\) We are aware of no other studies of universal interventions in the elementary grades that have investigated long-term effects on indicators of adult functioning.

The Seattle Social Development Project (SSDP) intervention was guided theoretically by the social development model.\(^19\)\(^20\) We sought to identify and develop methods of management and instruction that could be used by public school teachers and adult caretakers to set children on a positive developmental course by promoting opportunities for children's active involvement in classroom and family, developing children's skills for participation, and encouraging reinforcement from parents and teachers for children's effort and accomplishment. Two intervention conditions were examined: a full-intervention condition implemented throughout grades 1 through 6 and a late-intervention condition implemented only in grades 5 and 6.\(^21\)

Studies of the SSDP intervention have found significant effects in childhood and adolescence across outcomes.\(^22\)\(^23\) By age 21 years, the full-intervention group, compared with the control group, exhibited significantly better outcomes for education, employment, and mental health, as well as reduced crime, sexual risk behavior and disease, and early pregnancy.\(^24\)\(^25\) Some effects were found to be moderated by sex, race/ethnicity, or childhood poverty.\(^26\)\(^27\)\(^28\)\(^29\) Herein we examined the effects of the SSDP intervention at ages 24 and 27 years, 12 to 15 years after the intervention ended. The sex-balanced and racially/ethnically diverse sample enabled investigation of possible moderators of intervention effects.

Method

Sample and Design
The Figure shows the overall design of the study. Beginning in fall 1981, the intervention was initiated among first-grade students in classrooms randomly assigned to condition in 8 public schools serving high-crime areas in Seattle, Washington. Three hundred seventy-seven students who remained in or entered the 8 schools during grades 1 through 3 were observed prospectively to fifth grade. The study was then expanded to include 676 fifth-grade students in 10 additional schools, and all parents were asked for consent for their child to participate in the longitudinal follow-up study. Of the population of 1053 fifth-grade students in the 18 schools, parents of 808 children (77%) consented.

**Figure.**

Study flowchart.

Schools were assigned nonrandomly to the intervention or control condition in fall 1985. Thereafter, all fifth-grade students in each school received the intervention according to their school’s intervention assignment. This resulted in a nonrandomized controlled trial with 4 conditions. The full-intervention group consisted of those who received at least 1 semester of intervention in grades 1 through 4 and at least 1 semester of intervention in grades 5 and 6, with a mean of 4.13 years of intervention exposure. The late-intervention group consisted of those who received the intervention during grades 5 and 6 only, with a mean of 1.65 years of exposure. The control group received no intervention. A fourth group was offered parent training only during grades 5 and 6 and is not discussed in this article. Twenty-four participants could not be classified into any of these groups because they left participating schools before attending for at least 1 semester. After excluding these 24 participants, all analyses were based on intervention assignment. All phases of the study were approved by the Human Subjects Review Committee at the University of Washington, Seattle. Participants were informed about the nature of the interviews and provided consent before participation in the study at ages 24 and 27 years.

**BACKGROUND OF ANALYSIS SAMPLE**

Retention for analysis of the full-intervention, late-intervention, and control groups averaged 93% at both ages 24 and 27 years (Figure). There were equal numbers of female and male participants at each age. Racial/ethnic identification was 46% European American, 26% African American, 22% Asian American, and 6% Native American. As children, 56% of participants were eligible for the federal school lunch or breakfast program at some point in the fifth, sixth, or seventh grade, indicating low income status.

**ATTRITION AND INTERNAL VALIDITY**
At both ages 24 and 27 years, the overall distribution of participants in the intervention conditions did not differ significantly for those lost to attrition vs the analysis sample ($\chi^2 = 2.16, P = .34$ at age 24 years; and $\chi^2 = 1.41, P = .50$ at age 27 years); in addition, among those retained in the analysis sample, at both ages 24 and 27 years, the distribution of participants in the intervention conditions did not differ for sex ($\chi^2 = 0.38, P = .83$; and $\chi^2 = 0.59, P = .74$, respectively), race/ethnicity (white vs other: $\chi^2 = 0.13, P = .94$; and $\chi^2 = 0.02, P = .99$, respectively), or childhood poverty ($\chi^2 = 0.23, P = .89$; and $\chi^2 = 0.33, P = .85$, respectively).

Given the requirement that students who received the full intervention attended project schools at some point in grades 1 through 4 and in grades 5 and 6, whereas some students in the control group were added to the study at grade 5, it is important to rule out differences in residential stability, a potential threat to internal validity. Analyses comparing the full-intervention and control groups found no significant differences in mean number of years living in Seattle by grade 6 ($F = 0.61, P = .44$ at age 24 years; and $F = 1.83, P = .18$ at age 27 years), mean number of residences lived in from age 5 to 14 years ($F = 1.57, P = .21$; and $F = 1.56, P = .21$, respectively), percentage of single-parent homes during grade 5 ($\chi^2 = 0.11, P = .74$; and $\chi^2 = 0.02, P = .89$, respectively), or living in a disorganized neighborhood at age 16 years (eg, high crime or rundown housing: $\chi^2 = 0.47, P = .49$; and $\chi^2 = 0.13, P = .72$, respectively). Differential school or teacher receptivity to intervention is also an unlikely threat to internal validity. Teachers in 6 of 8 participating schools during grades 1 through 4 were randomly assigned to either intervention or control classrooms. At fifth grade, newly eligible schools were matched demographically to early experimental schools, and each agreed to serve as a control or late-intervention school depending on assignment. In addition, during the course of the intervention, the Seattle school district used mandatory busing to achieve racial/ethnic equality in schools, which substantially reduced the risk that outcomes observed reflected contextual or neighborhood differences, school demographic differences, or parent school-selection effects in the populations attending different schools.

An exception to the pattern of condition equivalence was the percentage of surveyed participants who reported at age 24 years that their mothers were aged 19 years or younger when they were born. Nine percent of the full-intervention condition compared with 21% of the control condition reported that their mothers were teenagers when they were born ($\chi^2 = 8.56, P < .01$). Having a teenaged mother was included as a covariate in all outcome analyses in this study.

**INTERVENTION**

The intervention is described elsewhere. Each year during grades 1 through 6, teachers in the intervention conditions received 5 days of in-service training in instructional methods. In addition, first-grade teachers received instruction in the use of a cognitive and social skills training curriculum, and during grade 6, a study consultant provided students with training in refusal skills. When children were in grades 1 through 3, parents were offered a 7-session curriculum in child behavior management skills and a 4-session curriculum in skills for supporting their children's academic development. During grades 5 and 6, parents of participants in intervention conditions were offered a 5-session curriculum designed to strengthen skills to reduce their children's risk of problem behaviors. Forty-three percent of parents of children eligible for the full intervention attended at least 1 parenting class during grades 1 through 3, and 29% of parents of children in eligible intervention conditions attended at least 1 class during grades 5 or 6, which indicates that the parent-training component had less reach than the teacher training and child social skills training components.

**MAIN OUTCOME MEASURES**

Primary outcome indices of success in each life domain were analyzed. The indices provide a limited set of comparisons for multivariate statistical tests of intervention effects. Analyses of primary outcome indices were followed up, where warranted, by analyses of specific behaviors, attitudes, and events to provide descriptive detail to the findings and to enable comparison with previous reports. Measures were participant self-reports of events in the last year at ages 24 and 27 years unless otherwise noted, and court records from age 10 to 24 years. Extreme values for open-ended numeric responses were limited to an appropriate maximum to limit the effects of outliers.

For school and work functioning, a median socioeconomic status (SES) attainment index was assessed by creating a dichotomous measure of those scoring at or above the median in completed education (attaining at least a high school diploma by age 24 or 27 years) or household income (at least $44,000 at age 24 years and at least $45,000 at age 27 years). Specific SES measures included cumulative education completed and income (divided by 1000 for analysis). Other measures included the degree to which students were integrated at school (eg, time in class and participation, coded 0 [poorly integrated] to 4 [well integrated]; mean reliability coefficient $\alpha$ at ages 24 and 27 years, $\alpha = 0.37$) and the degree of responsibility on the job for those who were employed (2 items coded 0 [low responsibility] to 4 [high responsibility]; mean $= .36$). (For purposes of comparison, measures at age 21 years were replicated as closely as possible. As a result, some reliability coefficients for measures corresponding to those examined at age 21 years are low.) Constructive engagement summed the average number of hours per week engaged in school or work (divided by 10 for analysis), and constructive self-efficacy concerned perceived opportunities for attaining a good education and a good job (coded 0 [low engagement] to 4 [high engagement]; $\alpha = .78$). Two additional descriptive measures, for which we did not expect intervention effects, were student status in the last year and employment status in the last month, ranging from nonstudent or unemployed (coded 0) to full-time student or employed (coded 4). To assess community involvement, a civic engagement index summed the mean number of hours per month of involvement in community groups and volunteer activities.
Mental health problems were assessed with the Diagnostic Interview Schedule (DSM-IV) to measure Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition) (DSM-IV) criteria. A disorder criterion index (coded 0-37; α = .95) summed across the total number of DSM-IV criteria met in the last year for a generalized anxiety disorder (GAD) criterion count (coded 0-6; α = .88), a social phobia criterion count (coded 0-5; α = .92), a posttraumatic stress disorder criterion count (coded 0-17; α = .98), and a major depressive episode criterion count (coded 0-9; α = .98). Specific criterion count measures for each of these disorders were also analyzed, including a separate measure of suicidal thoughts (coded 0-3; α = .67). In addition, a dichotomous disorder diagnosis index was computed to indicate those meeting criteria for DSM-IV diagnosis for any of the measured disorders (GAD, social phobia, posttraumatic stress disorder, or major depressive episode).

The lifetime sexually transmitted disease (STD) index was a dichotomous measure based on reports of having ever been diagnosed as having an STD. Specific measures of sexual risk behavior included the number of lifetime sex partners and, among those not in an exclusive relationship, the number of past-year sex partners and condom use in the last year (coded as percentage of time). We also report descriptive measures of having ever been or gotten someone pregnant and ever having or fathering a baby, for which we did not expect intervention effects. Lifetime measures were constructed accounting for affirmative responses in previous-year surveys.

A substance abuse and dependence criterion index was computed as the sum of DSM-IV criteria met for alcohol and illicit drug abuse and dependence disorders (coded 0-22; α = .86), and a substance abuse and dependence diagnosis index indicated those meeting criteria for an abuse or dependence diagnosis. Specific behaviors related to problem substance use included high variety of substance use (identifying those above the 90th percentile in number of different substances used in the last year) and the extent of substance use interference with life (coded 0 for no use to 4 “very much”). Any past-year substance use (tobacco, alcohol, or illicit drugs) is reported for descriptive purposes.

A past-year crime index was computed from self-reports of criminal acts other than driving violations or illicit drug use. Specific crime measures included high variety of crime, identifying those above the 90th percentile in number of different types of crimes committed, and measures of having sold drugs and having been arrested. In addition, official state and federal crime files were matched to SSDP participants (including survey nonrespondents) to assess the presence of a past-year court charge or a lifetime court charge for any noncriminal, misdemeanor, or felony charge through age 24 years.

DATA ANALYSIS

Because of the study design, the unit of intervention assignment consisted of the series of classrooms to which some students were assigned in grades 1 through 4 and the condition assignments of schools attended by all participants in grades 5 and 6. Of the 643 participants assigned to the control, late-intervention, and full-intervention conditions, more than 169 different classroom or school sequences were identified, consistent with the unit of intervention assignment. On average, only 3.80 participants experienced the same units of intervention within conditions. For this reason and to be consistent with previous reports, analyses were conducted at the individual level.

RESULTS

A multivariate analysis of covariance was conducted to assess overall intervention effects across multiple dependent variables, controlling for teenaged mother at birth. All 16 primary outcome indices (8 each at ages 24 and 27 years) were included in this analysis. Results showed a significant overall difference between the full-intervention and control groups with listwise deletion (Wilks Λ, Λ[16, 276] = 1.98; P = .01). This multivariate analysis of covariance was replicated across 5 data sets for which imputation procedures were used to account for missing data on some outcomes. Each analysis confirmed an overall group difference (Wilks Λ, Λ[16, 109] = 1.84-1.94; P = .03-.02). These results indicate a significant multivariate effect and provide overall control for type I error rate in the significant univariate findings. Multivariate analyses of covariance comparing the late-intervention group with the control group across the 16 outcomes were not significant (Wilks Λ, Λ[16, 425] = 0.69-0.75; P = .81-.74).

SCHOOL, WORK, AND COMMUNITY

Participants in the full-intervention group were significantly more likely than those in the control group to be at or above the median in SES (educational attainment or household income) by age 27 years (93% vs 84% in the control group) (Table). The differences in SES attainment were similar at age 24 years, though not significant. Specific comparisons in this domain found that those in the full-intervention groups were marginally more likely to have continued their education beyond high school, with 34% having completed an associate degree by age 27 years compared with 22% in the control group (P < .06). Nonsignificant trends in completion of a bachelor’s degree and in household and earned income also favored the full-intervention group.

Table. Comparison of Control, Late-, and Full-Intervention Groups Across Adult Outcomes at Ages 24 and 27 Years
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Age 24 Years</th>
<th>Age 27 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>School, work, and community</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median SES attainment index</td>
<td>566 (0.86) (0.35) (0.26)</td>
<td>595 (0.84) (0.37) (0.25)</td>
</tr>
<tr>
<td>Civic engagement index</td>
<td>598 (9.22) (10.93) (14.91)</td>
<td>598 (9.55) (15.60) (14.59)</td>
</tr>
<tr>
<td>Specific SES measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completed GED&lt;sup&gt;3&lt;/sup&gt;</td>
<td>596 (0.90) (0.27) (0.28)</td>
<td>597 (0.91) (0.29) (0.28)</td>
</tr>
<tr>
<td>Completed high school&lt;sup&gt;2&lt;/sup&gt;</td>
<td>596 (0.75) (0.41) (0.39)</td>
<td>597 (0.77) (0.42) (0.39)</td>
</tr>
<tr>
<td>Completed associate's degree&lt;sup&gt;2&lt;/sup&gt;</td>
<td>596 (0.22) (0.41) (0.47)</td>
<td>596 (0.22) (0.42) (0.48)</td>
</tr>
<tr>
<td>Completed bachelor's degree&lt;sup&gt;2&lt;/sup&gt;</td>
<td>596 (0.13) (0.34) (0.40)</td>
<td>596 (0.14) (0.35) (0.40)</td>
</tr>
<tr>
<td>Household income, $&lt;sup&gt;2&lt;/sup&gt; (in thousands)</td>
<td>521 (49.92) (36.97) (36.83)</td>
<td>577 (54.12) (47.45) (42.27)</td>
</tr>
<tr>
<td>Earned income, $&lt;sup&gt;2&lt;/sup&gt; (in thousands)</td>
<td>563 (19.65) (17.75) (18.66)</td>
<td>590 (25.03) (20.56) (20.09)</td>
</tr>
<tr>
<td>Specific behaviors and attitudes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated at school</td>
<td>165 (1.58) (0.59) (0.74)</td>
<td>135 (1.49) (0.59) (0.59)</td>
</tr>
<tr>
<td>Responsibility on job</td>
<td>561 (2.43) (1.06) (1.09)</td>
<td>546 (2.48) (1.08) (1.17)</td>
</tr>
<tr>
<td>Constructive engagement</td>
<td>594 (2.74) (1.55) (1.46)</td>
<td>595 (3.16) (1.26) (1.33)</td>
</tr>
<tr>
<td>Constructive self-efficacy</td>
<td>597 (3.19) (0.64) (0.58)</td>
<td>598 (3.07) (0.67) (0.73)</td>
</tr>
<tr>
<td>Community groups, h/mo</td>
<td>598 (5.96) (9.97) (12.76)</td>
<td>596 (7.07) (11.11) (12.69)</td>
</tr>
<tr>
<td>Volunteer activities, h/mo</td>
<td>597 (4.49) (9.62) (12.08)</td>
<td>596 (3.99) (8.79) (10.04)</td>
</tr>
<tr>
<td>Descriptive measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student status</td>
<td>594 (0.79) (1.23) (1.28)</td>
<td>598 (0.61) (1.14) (1.09)</td>
</tr>
<tr>
<td>Employment status</td>
<td>597 (3.53) (1.03) (1.03)</td>
<td>593 (3.45) (1.16) (1.19)</td>
</tr>
<tr>
<td>Mental health</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disorder criterion index</td>
<td>596 (4.45) (7.89) (5.66)</td>
<td>598 (6.48) (7.97) (5.20)</td>
</tr>
<tr>
<td>Disorder diagnosis index&lt;sup&gt;2&lt;/sup&gt;</td>
<td>596 (0.27) (0.45) (0.39)</td>
<td>598 (0.26) (0.44) (0.41)</td>
</tr>
<tr>
<td>Specific disorders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAD criterion count</td>
<td>595 (0.76) (1.87) (1.71)</td>
<td>598 (2.20) (2.04) (1.95)</td>
</tr>
<tr>
<td>Social phobia criterion count</td>
<td>596 (0.27) (1.07) (1.08)</td>
<td>598 (1.24) (1.73) (1.58)</td>
</tr>
<tr>
<td>PTSD criterion count</td>
<td>594 (2.05) (4.52) (3.70)</td>
<td>598 (1.22) (3.95) (3.17)</td>
</tr>
<tr>
<td>MDE criterion count</td>
<td>595 (1.39) (5.13) (3.11)</td>
<td>596 (1.82) (3.33) (3.00)</td>
</tr>
<tr>
<td>Suicidal thoughts</td>
<td>595 (0.21) (0.65) (0.68)</td>
<td>597 (0.17) (0.47) (0.43)</td>
</tr>
<tr>
<td>GAD diagnosis&lt;sup&gt;2&lt;/sup&gt;</td>
<td>595 (0.11) (0.31) (0.25)</td>
<td>597 (0.07) (0.26) (0.26)</td>
</tr>
<tr>
<td>Social phobia diagnosis&lt;sup&gt;2&lt;/sup&gt;</td>
<td>596 (0.03) (0.17) (0.15)</td>
<td>598 (0.08) (0.27) (0.21)</td>
</tr>
<tr>
<td>PTSD diagnosis&lt;sup&gt;2&lt;/sup&gt;</td>
<td>594 (0.12) (0.33) (0.26)</td>
<td>598 (0.08) (0.28) (0.20)</td>
</tr>
<tr>
<td>MDE diagnosis&lt;sup&gt;2&lt;/sup&gt;</td>
<td>595 (0.16) (0.35) (0.31)</td>
<td>598 (0.15) (0.36) (0.36)</td>
</tr>
</tbody>
</table>

<sup>2</sup> Mean or Prevalence (SD)  
<sup>3</sup> P Value<sup>3</sup>  
<sup>4</sup> Mean or Prevalence (SD)  
<sup>5</sup> P Value<sup>5</sup>
The civic engagement index indicated marginally more community involvement and volunteerism in the full-intervention group at ages 24 and 27 years compared with the control group. Both intervention groups reported significantly more participation in community groups at age 24 years compared with the control group.

Across outcomes related to school, work, and community at ages 24 and 27 years for which we had directional hypotheses, 27 of 28 outcomes analyzed showed directional patterns between the full-intervention and control conditions consistent with prediction, and both primary outcome indices showed positive full-intervention effects significant beyond $P < .08$ by age 27 years. Twenty-two of these 28 outcomes demonstrated a pattern consistent with a dose effect, with the late-treatment group falling between the full-treatment and control groups.

MENTAL HEALTH

Analyses found significantly fewer symptoms of mental health disorders on the disorder criterion index in the full-treatment condition compared with the control group at both ages 24 and 27 years. Analyses also found significantly lower prevalence of those meeting criteria for at least 1 of 4 DSM-IV diagnoses on the disorder diagnosis index in the full-treatment group compared with the control group at age 27 years and a nonsignificant trend toward lower prevalence at age 24 years. Measures of specific disorder criterion counts indicated at least marginally significant ($P < .10$) reductions in the number of GAD, social phobia, and posttraumatic stress disorder criteria met at age 24 years and of GAD, posttraumatic stress disorder, and major depressive episode criteria met, as well as significantly fewer suicidal thoughts, at age 27 years. Across all 22 mental health outcomes examined, reported problems were lower in magnitude in the full-intervention group compared with the control group. Participants in the late-intervention group reported levels between or equal to those of the other groups for all but 3 of the 22 mental health outcomes, consistent with a dose effect.

SEXUAL BEHAVIOR AND PARENTHOOD

The lifetime STD index showed significantly lower prevalence of having ever been diagnosed with an STD at ages 24 and 27 years for the full-intervention group compared with the control group. No intervention effects on specific sexual behaviors were found.

SUBSTANCE USE AND CRIME

No intervention effects on either of the substance abuse and dependence indices were found in young adulthood, and no significant effects were observed for the specific measures of high variety of substance use or substance use interference with life. A marginally significant trend ($P < .09$), counter to hypotheses, for the last-year crime index at age 27 years indicated a higher prevalence of having committed a crime among participants in the full-intervention group compared with the control group. This difference stemmed primarily from nonsignificant but higher rates in the full-intervention group of minor theft (having taken anything worth less than $50; 9% vs 4% in the control group) and drug selling (8% vs 3%, respectively) at age 27 years. This was the only finding counter to hypotheses that approached significance across all of the outcomes examined in the Table. No significant intervention effects were found for any specific measures related to crime.

INTERACTIONS WITH SEX, POVERTY, AND RACE/ETHNICITY

Adding interaction terms to the regression models revealed 2 significant effects of the full intervention specific to male subjects. Male subjects in the full-treatment group were significantly more likely to meet median education or income criteria at age 24 years as assessed using the SES attainment index compared with their counterparts in the control group (97% [SD, 0.17] vs 85% [SD, 0.36], respectively; $P = .03$). In addition, male subjects in the full-treatment group were significantly more constructively engaged at age 24 years compared with male subjects in the control group (mean [SD] 29.9 [13.9] h/wk of engagement in school or work vs 24.4 [16.4] h/wk, respectively; $P = .02$). Neither of these outcomes showed significant intervention differences for female subjects.

Two significant intervention effects were also found specific to participants from a background of childhood poverty. Those in the full-treatment group from a background of childhood poverty reported a mean (SD) of almost 9.5 (13.2) hours per month of involvement in community groups at age 27 years compared with 4.9 (9.2) hours per month for those in the control group ($P = .004$). Participants from childhood poverty in the full-intervention group were also significantly less likely to meet sufficient criteria for a diagnosis of GAD at age 24 years compared with those in the control group (3% [SD, 0.16] vs 14% [SD, 0.35], respectively; $P = .02$).

After controlling for childhood poverty, race/ethnicity also significantly interacted with the intervention for some outcomes. Responsibility on the job at age 24 years was significantly higher among white participants in the full-treatment group compared with whites in the control group (mean = 2.78 [SD, 1.10] and mean = 2.34 [SD, 1.06], respectively; $P = .02$). African Americans in the full-intervention condition compared with their control counterparts reported significantly higher household income at age 27 years (mean = $55,594 [SD, $45,966] and mean = $35,288 [SD = $35,810], respectively; $P = .046$), fewer GAD criteria met at age 27 years (mean = 0.94 [SD, 1.39] and mean = 2.51 [SD, 1.99], respectively; $P < .001$), fewer past-year sex partners at age 27 years (mean = 1.39 [SD, 1.65] and mean = 2.89
[SD, 2.78], respectively; \( P = .03 \), and reduced prevalence of lifetime STD diagnosis at age 24 years (12% [SD, 0.33] in the full-intervention condition and 55% [SD, 0.50] in the control condition; \( P = .001 \)) and at age 27 years (16% [SD, 0.37] in the full-intervention condition and 61% [SD, 0.49] in the control condition; \( P = .001 \)). These results suggest that the significant main effects of the full intervention on STD diagnoses at ages 24 and 27 years were primarily because of the reductions in STD diagnoses among African Americans in the full-intervention group.

**COMMENT**

The SSDP intervention in the elementary grades showed a significant overall effect of the full intervention in a multivariate analysis of 16 outcome indices at ages 24 and 27 years, 12 and 15 years after the intervention ended. Summary indices revealed significantly better SES, mental health, and sexual health by age 27 years in those assigned to the SSDP full-intervention condition compared with those in the control group. In contrast, effects of the intervention on crime and substance use were not found at ages 24 and 27 years.

In line with previous analyses,32 33 the rank of means and prevalence rates for the late-intervention group compared with the full-intervention and control groups suggests a dose effect for the domains most influenced by the intervention. That is, although differences were not significant, the late-intervention group reported better outcomes than the control group across the preponderance of measures for which there were directional hypotheses related to school, work, mental health, or community, but worse outcomes compared with the full-intervention group across these same measures.

Our findings indicate that a theory-based intervention that improved parenting practices, children’s social competence, and classroom management and instruction during the elementary grades influenced some, though not all, indices of adult functioning in individuals in their mid- to late 20s. The elementary grade intervention was associated with greater accomplishment and engagement in school, work, and community and fewer mental health problems by ages 24 and 27 years. Effects of the SSDP intervention on mental health outcomes are particularly noteworthy at this stage of young adulthood, given the potentially debilitating consequences of depression and anxiety.36 39

Findings should be considered in the context of the design and analysis approach, which was based on condition assignment rather than attendance or implementation, using controlled 2-tailed statistical tests and an omnibus test of multiple outcomes. Strengths of this study include the ability to detect relatively small effect sizes, comprehensive assessment capturing multiple outcome domains, delivery of the intervention package universally to all students in intervention classrooms, and the ability to investigate possible moderating effects of sex, race/ethnicity, and childhood poverty. As in previous reports,33 there was little evidence of serious threats to internal validity. In addition, procedures for assigning classrooms (in the early experiment) and schools to conditions likely guarded against possible effects of differential school or teacher receptivity to intervention, and mandatory busing provided further protection against confounding neighborhood or school demographic differences or parent school-selection effects. An exception to the general pattern of group equivalence at the start of the longitudinal study was a significantly higher percentage of participants in the control group who reported having had a teenaged mother at birth. This difference was controlled for statistically throughout all analyses.

Limitations should be noted. The study was quasi-experimental and geographically limited. It relied heavily on self-reported data from study participants. Effects of the intervention on school, work, and mental health were observed, but no significant effects on substance use or crime were observed at these ages.

Intervention effects reported herein indicate that universal intervention during the elementary grades to improve the management and instructional skills of urban public elementary schoolteachers, strengthen parenting practices in multiracial/multiethnic urban families, and ensure that children have the skills to participate in the social and academic life of elementary school can positively affect attainment, functioning, and mental health in young adulthood.

**ARTICLE INFORMATION**

**Correspondence:** J. David Hawkins, PhD, Social Development Research Group, School of Social Work, University of Washington, 9725 Third Ave NE, Ste 401, Seattle, WA 98115 (jdh@u.washington.edu).

**Accepted for Publication:** April 23, 2008.

**Author Contributions:** Dr Hawkins had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. **Study concept and design:** Hawkins, Kosterman, Catalano, and Hill. **Acquisition of data:** Hawkins and Hill. **Analysis and interpretation of data:** Kosterman, Catalano, Hill, and Abbott. **Drafting of the manuscript:** Kosterman. **Critical revision of the manuscript for important intellectual content:** Hawkins, Kosterman, Catalano, Hill, and Abbott. **Statistical analysis:** Kosterman and Abbott. **Obtained funding:** Hawkins, Kosterman, Catalano, and Hill. **Administrative, technical, and material support:** Hill. **Study supervision:** Hawkins.

**Financial Disclosure:** Dr Catalano receives reimbursement for being a member of the board of Channing Bete Co, distributor of Supporting School Success and Guiding Good Choices.
Funding/Support: This study was supported by grants 1R01DA09679-11 and 9R01DA021426-08 from the National Institute on Drug Abuse, R24MH56587-06 from the National Institute of Mental Health, and 21548 from the Robert Wood Johnson Foundation.

Additional Information: Supporting School Success and Guiding Good Choices were tested in the study that produced the data set used in this article.

REFERENCES

3 Bazargan MCalderon JL Heslin KC Mentes CShaheen MAHdout JBaker RA A profile of chronic mental and physical conditions among African-American and Latino children in urban public housing. Ethn Dis 2005 Autumn;15 ((4 Suppl 5)) S5-3- 9 PubMed
4 Blanchett WJMumford VBeachum F Urban school failure and disproportionality in a post-Brown era: benign neglect of the constitutional rights of students of color. Remedial Special Educ2005;26 (2) 70- 81 Link to Article
6 Oh JH Social disorganizations and crime rates in US central cities: toward an explanation of urban economic change. Soc Sci J 2005;42 (4) 569- 582 Link to Article
8 Galaif EREnewcomb MDCarmona JV Prospective relationships between drug problems and work adjustment in a community sample of adults. J Appl Psychol 2001;86 (2) 337- 350 PubMed Link to Article
10 Rindfuss RR Cooksey ECSutterlin RL Young adult occupational achievement: early expectations versus behavioral reality. Work Occupations 1999;26 (2) 220- 263 Link to Article
13 Kessler RCMcGonagle KAZhao S et al. Lifetime and 12-month prevalence of DSM-III-Rpsychiatric disorders in the United States. Arch Gen Psychiatry 1994;51 (1) 8- 19 PubMed Link to Article
14 Eaton WWAnthony JC Gallo J et al. Natural history of Diagnostic Interview Schedule/DSM-IV major depression: the Baltimore Epidemiologic Catchment Area follow-up. Arch Gen Psychiatry1997;54 (11) 993- 999 PubMed Link to Article
15 Binson D Dolcini MMPollack LMCatania JA Data from the National AIDS Behavioral Surveys: IV, multiple sexual partners among young adults in high-risk cities. Fam Plann Perspect1993;25 (6) 268- 272 PubMed Link to Article
17 Bachman JGO’Malley PM Schulenberg JEO’Johnston LDL Bryant AL Merline AC The Decline of Substance Use in Young Adulthood: Changes in Social Activities, Roles, and Beliefs. Mahwah, NJ Lawrence Erlbaum Associates2002;
19 Elliott DS Serious violent offenders: onset, developmental course, and termination: American Society of Criminology 1993 Presidential Address. Criminology 1994;32 (1) 1- 22 Link to Article


35 Hawkins JD, Catalano RF, Jones GF, Fine DN. Delinquency prevention through parent training: results and issues from work in progress. Wilson JQ, Loury GC. From *Children to Citizens Families, Schools, and Delinquency Prevention*. New York, NY: Springer-Verlag 1987; 186-204


37 Cummings C. Managing to Teach. Edmonds, WA: Teaching Inc 1983;


44 Hawkins JD, Catalano RF. Preparing for School Success. South Dearfield, MA: Channing Bete Co 2003;

45 Hawkins JD, Catalano RF. Guiding Good Choices. South Dearfield, MA: Channing Bete Co 2003;

49 Jaffee SRMoffitt TECaspi AFombonne EPoulton RMartin J Differences in early childhood risk factors for juvenile-onset and adult-onset depression. *Arch Gen Psychiatry* 2002;59 (3) 215- 222 PubMed Link to Article
56 Eaton WW The Sociology of Mental Disorders. 3rd ed. Westport, CT Praeger 2001;
57 Wells JCTien AYGarrison REaton WW Risk factors for the incidence of social phobia as determined by the Diagnostic Interview Schedule according to DSM-III in a population-based study. *Acta Psychiatr Scand* 1994;90 (2) 84- 90 PubMed Link to Article
58 Booth BMZhang MRost KMCaldary JASmith LGSmith GR Measuring outcomes and costs for major depression. *Psychopharmacol Bull* 1997;33 (4) 653- 658 PubMed
59 Davila JBeck JG Is social anxiety associated with impairment in close relationships? a preliminary investigation. *Behav Ther* 2002;33 (3) 427- 447 Link to Article
The Impact of Enhancing Students’ Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions

Joseph A. Durlak¹, Roger P. Weissberg², Allison B. Dymnicki³, Rebecca D. Taylor⁴, Kriston B. Schellinger⁴

This article presents findings from a meta-analysis of 213 school-based, universal social and emotional learning (SEL) programs involving 270,034 kindergarten through high school students. Compared to controls, SEL participants demonstrated significantly improved social and emotional skills, attitudes, behavior, and academic performance that reflected an 11-percentile-point gain in achievement. School teaching staff successfully conducted SEL programs. The use of 4 recommended practices for developing skills and the presence of implementation problems moderated program outcomes. The findings add to the growing empirical evidence regarding the positive impact of SEL programs. Policy makers, educators, and the public can contribute to healthy development of children by supporting the incorporation of evidence-based SEL programming into standard educational practice.

Teaching and learning in schools have strong social, emotional, and academic components (Zins, Weissberg, Wang, & Walberg, 2004). Students typically do not learn alone but rather in collaboration with their teachers, in the company of their peers, and with the encouragement of their families. Emotions can facilitate or impede children’s academic engagement, work ethic, commitment, and ultimate school success. Because relationships and emotional processes affect how and what we learn, schools and families must effectively address these aspects of the educational process for the benefit of all students (Elias et al., 1997).

A key challenge for 21st-century schools involves serving culturally diverse students with varied abilities and motivations for learning (Learning First Alliance, 2001). Unfortunately, many students lack social-emotional competencies and become less connected to school as they progress from elementary to middle to high school, and this lack of connection negatively affects their academic performance, behavior, and health (Blum & Libbey, 2004). In a national sample of 148,189 sixth to twelfth graders, only 29%–45% of surveyed students reported that they had social competencies such as empathy, decision making, and conflict resolution skills, and only 29% indicated that their school provided a caring, encouraging environment (Benson, 2006). By high school as many as 40%–60% of students become chronically disengaged from school (Klem &Connell, 2004). Furthermore, approximately 30% of high school students engage in multiple high-risk behaviors (e.g., substance use, sex, violence, depression, attempted suicide) that interfere with school performance and jeopardize their potential for life success (Dryfoos, 1997; Eaton et al., 2008).

There is broad agreement among educators, policy makers, and the public that educational systems should graduate students who are proficient in core academic subjects, able to work well with others from diverse backgrounds in socially and emotionally skilled ways, practice healthy behaviors, and behave responsibly and respectfully (Association for Supervision and Curriculum Development, 2007; Greenberg et al., 2003). In other words, schools have an important role to play in raising healthy children by fostering not only their cognitive development but also their social and emotional development. Yet schools have limited resources to address all of these areas and are experiencing intense pressures to enhance academic performance. Given time constraints and competing demands, educators must prioritize and effectively implement evidence-based approaches that produce multiple benefits.

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

What Is Social and Emotional Learning?

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

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

Recent Relevant Research Reviews

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

The Current Meta-Analysis: Research Questions and Hypotheses

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

The proliferation of new competence-promotion approaches led to several important research questions about school-based interventions to foster students’ social and emotional development. For example, what outcomes are achieved by interventions that attempt to enhance children’s emotional and social skills? Can SEL interventions promote positive outcomes and prevent future problems? Can programs be successfully conducted in the school setting by existing school personnel? What variables moderate the impact of school-based SEL programs? Next, we address these questions and offer hypotheses about expected findings.
The findings from several individual studies and narrative reviews indicate that SEL programs are associated with positive results such as improved attitudes about the self and others, increased prosocial behavior, lower levels of problem behaviors and emotional distress, and improved academic performance (Catalano et al., 2002; Greenberg et al., 2003; Zins et al., 2004). Thus, our first hypothesis was that our meta-analysis of school-based SEL programs would yield significant positive mean effects across a variety of skill, attitudinal, behavioral, and academic outcomes (Hypothesis 1). Ultimately, interventions are unlikely to have much practical utility or gain widespread acceptance unless they are effective under real-world conditions. Thus, we investigated whether SEL programs can be incorporated into routine educational practice; that is, can they be successfully delivered by existing school staff during the regular school day? In our analyses, we separated interventions conducted by regular school staff and those administered by nonschool personnel (e.g., university researchers, outside consultants). We predicted that programs conducted by classroom teachers and other school staff would produce significant outcomes (Hypothesis 2).

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

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

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

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

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

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

Method

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

**Inclusion Criteria**

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

**Exclusion Criteria**

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

**Dealing With Multiple Cohorts or Multiple Publications on the Same Cohort**

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

**Independent Variable: Intervention Formats**

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

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

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

Potential Moderators of Outcome: SAFE and Implementation

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

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

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

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

Social and emotional skills. This category includes evaluations of different types of cognitive, affective, and social skills related to such areas as identifying emotions from social cues, goal setting, perspective taking, interpersonal problem solving, conflict resolution, and decision making. Skill assessments could be based on the reports from the student, a teacher, a parent, or an independent rater. However, all the outcomes in this category reflected skill acquisition or performance assessed in test situations or structured tasks (e.g., interviews, role plays, or questionnaires). In contrast, teacher ratings of students’ behaviors manifested in daily situations (e.g., a student’s ability to control anger or work well with others) were placed in the positive social behavior category below.

Attitudes toward self and others. This category combines positive attitudes about the self, school, and social topics. It included self-perceptions (e.g., self-esteem, self-concept, and self-efficacy), school bonding (e.g., attitudes toward school and teachers), and conventional (i.e., prosocial) beliefs about violence, helping others, social justice, and drug use. All the outcomes in this category were based on student self-reports. We combined these three outcomes to avoid extremely small cell sizes for subsequent analyses.

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

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

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

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

**Coding Reliability**

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

**Calculation of Effects and General Analytic Strategies**

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

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

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

**Results**

**Descriptive Characteristics of Reviewed Studies**

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

<table>
<thead>
<tr>
<th>General publication features</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955–1979</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>1980–1989</td>
<td>35</td>
<td>16</td>
</tr>
<tr>
<td>1990–1999</td>
<td>83</td>
<td>39</td>
</tr>
<tr>
<td>2000–2007</td>
<td>77</td>
<td>36</td>
</tr>
<tr>
<td>Source of report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Published article/books</td>
<td>172</td>
<td>81</td>
</tr>
<tr>
<td>Unpublished reports</td>
<td>41</td>
<td>19</td>
</tr>
<tr>
<td>Methodological features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>99</td>
<td>47</td>
</tr>
<tr>
<td>No</td>
<td>114</td>
<td>53</td>
</tr>
<tr>
<td>Mean percent of attrition</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Implementation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not reported on</td>
<td>91</td>
<td>43</td>
</tr>
<tr>
<td>No significant problems reported</td>
<td>74</td>
<td>35</td>
</tr>
<tr>
<td>Significant problems reported</td>
<td>48</td>
<td>22</td>
</tr>
<tr>
<td>Use of reliable outcome measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>550</td>
<td>76</td>
</tr>
<tr>
<td>No</td>
<td>176</td>
<td>24</td>
</tr>
<tr>
<td>Use of valid outcome measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>369</td>
<td>51</td>
</tr>
<tr>
<td>No</td>
<td>357</td>
<td>49</td>
</tr>
<tr>
<td>Source of outcome data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child</td>
<td>382</td>
<td>53</td>
</tr>
<tr>
<td>Other (parent, teacher, observer, school records)</td>
<td>422</td>
<td>47</td>
</tr>
</tbody>
</table>

Participant features

Educational level of participants

<table>
<thead>
<tr>
<th>Educational level of participants</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary school (Grades K–5)</td>
<td>120</td>
<td>56</td>
</tr>
<tr>
<td>Middle school (Grades 6–8)</td>
<td>66</td>
<td>31</td>
</tr>
<tr>
<td>High school (Grades 9–12)</td>
<td>27</td>
<td>13</td>
</tr>
</tbody>
</table>

Intervention features

Intervention format

<table>
<thead>
<tr>
<th>Intervention format</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class by Teacher</td>
<td>114</td>
<td>53</td>
</tr>
<tr>
<td>Class by Nonschool Personnel</td>
<td>44</td>
<td>21</td>
</tr>
<tr>
<td>Multicomponent</td>
<td>55</td>
<td>26</td>
</tr>
</tbody>
</table>
**General publication features**

<table>
<thead>
<tr>
<th>Use of recommended training procedures</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention rated as SAFE</td>
<td>176</td>
<td>83</td>
</tr>
<tr>
<td>Intervention not rated as SAFE</td>
<td>37</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of sessions</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean number of sessions</td>
<td>40.8</td>
<td></td>
</tr>
<tr>
<td>Median number of sessions</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Locale of intervention</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>186</td>
<td>87</td>
</tr>
<tr>
<td>Outside the United States</td>
<td>27</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>General area of school</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>99</td>
<td>47</td>
</tr>
<tr>
<td>Suburban</td>
<td>35</td>
<td>16</td>
</tr>
<tr>
<td>Rural</td>
<td>31</td>
<td>15</td>
</tr>
<tr>
<td>Combination of areas</td>
<td>30</td>
<td>14</td>
</tr>
<tr>
<td>Did not report</td>
<td>18</td>
<td>8</td>
</tr>
</tbody>
</table>

**Table 1. Descriptive Characteristics of 213 School-Based Universal Interventions With Outcomes at Post**

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

The grand study-level mean for all 213 interventions was 0.30 (CI = 0.26–0.33), which was statistically significant from zero. The $Q$ value of 2,453 was significant ($p \leq .001$) and the $I^2$ was high (91%), indicating substantial heterogeneity among studies and suggesting the existence of one or more variables that might moderate outcomes. **Table 2** presents the mean effects and their 95% confidence intervals obtained at post across all reviewed programs in each outcome category. All six means (range = 0.22 to 0.57) are significantly greater than zero and confirm our first hypothesis. Results (based on 35–112 interventions depending on the outcome category) indicated that, compared to controls, students demonstrated enhanced SEL skills, attitudes, and positive social behaviors following intervention, and also demonstrated fewer conduct problems and had lower levels of emotional distress. Especially noteworthy from an educational policy perspective, academic performance was significantly improved. The overall mean effect did not differ significantly for test scores and grades (mean ESs = 0.27 and 0.33, respectively). Although only a subset of studies collected information on academic performance, these investigations contained large sample sizes and involved a total of 135,396 students.

**Outcomes**

<table>
<thead>
<tr>
<th>SEL skills</th>
<th>Attitudes</th>
<th>Positive social behavior</th>
<th>Conduct problems</th>
<th>Emotional distress</th>
<th>Academic performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.57*</td>
<td>0.23*</td>
<td>0.24*</td>
<td>0.22*</td>
<td>0.24*</td>
</tr>
</tbody>
</table>

1. *$p \leq .05$.

**Group**

<table>
<thead>
<tr>
<th>Total sample</th>
<th>CI 0.48 to 0.62</th>
<th>CI 0.16 to 0.30</th>
<th>CI 0.16 to 0.29</th>
<th>CI 0.14 to 0.35</th>
<th>CI 0.15 to 0.39</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>68</td>
<td>106</td>
<td>86</td>
<td>112</td>
<td>49</td>
</tr>
<tr>
<td>ES</td>
<td>0.62*</td>
<td>0.23*</td>
<td>0.26*</td>
<td>0.20*</td>
<td>0.25*</td>
</tr>
</tbody>
</table>

**Class by Teacher**

<table>
<thead>
<tr>
<th>Class by Teacher</th>
<th>CI 0.41 to 0.82</th>
<th>CI 0.17 to 0.29</th>
<th>CI 0.15 to 0.38</th>
<th>CI 0.12 to 0.29</th>
<th>CI 0.08 to 0.43</th>
<th>CI 0.16 to 0.52</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>40</td>
<td>59</td>
<td>59</td>
<td>53</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 2. Mean Effects and .05 Confidence Intervals at Post for Total Sample and Each Intervention Format

### Follow-Up Effects

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

### School Staff Can Conduct Successful SEL Programs

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

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

### What Moderates Program Outcomes?

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

<table>
<thead>
<tr>
<th>Skills</th>
<th>Attitudes</th>
<th>Social behavior</th>
<th>Conduct problems</th>
<th>Emotional distress</th>
<th>Academic performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$ 63</td>
<td>80</td>
<td>73</td>
<td>88</td>
<td>33</td>
<td>24</td>
</tr>
<tr>
<td>ES 0.01</td>
<td>0.16*</td>
<td>0.02</td>
<td>0.16*</td>
<td>0.18</td>
<td>0.26*</td>
</tr>
</tbody>
</table>

Did not meet SAFE criteria

<table>
<thead>
<tr>
<th>CI</th>
<th>−0.57 to 0.6007</th>
<th>0.07 to 0.25</th>
<th>0.07 to 0.16*</th>
<th>−0.02 to 0.37</th>
<th>0.11 to 0.40</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>5</td>
<td>26</td>
<td>13</td>
<td>24</td>
<td>16</td>
</tr>
</tbody>
</table>

Implementation

<table>
<thead>
<tr>
<th>ES</th>
<th>0.58*</th>
<th>0.17*</th>
<th>0.32*</th>
<th>0.24*</th>
<th>0.21*</th>
<th>0.31*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI</td>
<td>0.33 to 0.83</td>
<td>0.09 to 0.24017 to 0.47</td>
<td>0.13 to 0.34</td>
<td>0.04 to 0.38</td>
<td>0.18 to 0.45</td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>29</td>
<td>46</td>
<td>33</td>
<td>35</td>
<td>22</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ES</th>
<th>0.86*</th>
<th>0.29</th>
<th>0.31*</th>
<th>0.27*</th>
<th>0.35*</th>
<th>0.33*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI</td>
<td>0.59 to 1.12</td>
<td>0.21 to 0.37017 to 0.45</td>
<td>0.18 to 0.36</td>
<td>0.16 to 0.54</td>
<td>0.20 to 0.46</td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>26</td>
<td>36</td>
<td>34</td>
<td>45</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

No problems

<table>
<thead>
<tr>
<th>ES</th>
<th>0.35</th>
<th>0.19*</th>
<th>0.01</th>
<th>0.15*</th>
<th>0.15</th>
<th>0.14</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI</td>
<td>−0.01 to 0.710.10 to 0.28 to 0.19</td>
<td>−0.08 to 0.38</td>
<td>−0.01 to 0.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td>13</td>
<td>24</td>
<td>19</td>
<td>32</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3. Findings for Moderator Analyses at Post by Outcome Category for Total Sample

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

<table>
<thead>
<tr>
<th>Grouping variable</th>
<th>Values across all outcomes</th>
<th>Values within each outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Q$</td>
<td>$I^2$</td>
</tr>
<tr>
<td></td>
<td>Between</td>
<td>Within</td>
</tr>
<tr>
<td>Skills</td>
<td>Attitudes</td>
<td>Social behavior</td>
</tr>
<tr>
<td>SAFE practices</td>
<td>41.6*</td>
<td>530.2*</td>
</tr>
<tr>
<td>For each outcome</td>
<td>$Q$ within</td>
<td>193.9*</td>
</tr>
<tr>
<td>$I^2$ within</td>
<td>65</td>
<td>0</td>
</tr>
<tr>
<td>Implementation</td>
<td>5.3*</td>
<td>74.8</td>
</tr>
<tr>
<td>For each outcome</td>
<td>$Q$ within</td>
<td>74.8</td>
</tr>
<tr>
<td>$I^2$ within</td>
<td>12</td>
<td>14</td>
</tr>
</tbody>
</table>

1. $p \leq .05$.
Table 4. \( Q \) Statistics and \( I^2 \) Values (in Percent) for Study Groupings for Moderator Analyses

The data in Table 4 support the notion that both SAFE and implementation problems moderate SEL outcomes. For example, based on \( I^2 \) values, initially dividing ESs according to the six outcomes does produce the preferred low overall degree of within-group heterogeneity (15%) and high between-group heterogeneity (88%); for two specific outcomes, however, there is a mild (positive social behaviors, 32%) to moderately high (skills, 65%) degree of within-group heterogeneity. When the studies are further divided by SAFE practices or by implementation problems, the overall within-group variability remains low (12% and 13%, respectively), the within-group heterogeneity for both skills and social behaviors is no longer significant according to \( Q \) statistics, \( I^2 \) values drop to low levels (≤ 15%) and remain low for the other outcomes as well, and heterogeneity levels attributed to differences between groups are high or moderate (\( I^2 \) values of 79% and 63% for SAFE and implementation, respectively). In other words, the use of all four SAFE practices and reported implementation problems to subdivide groups provided a good fit for the obtained data.

These latter findings are consistent with the mean differences between groups on many outcomes for the SAFE and implementation data presented in Table 3. SAFE and implementation problems were not significantly correlated (\( r = -.07 \)). However, it was not possible to explore their potential interactions as moderators because only 57% of the studies monitored implementation and subdividing the studies created extremely small cell sizes that would not support reliable results.

Inspection of the distribution of the moderator variables in the different cells in Table 3 indicated that SAFE practices and implementation problems were more common for some intervention formats. Compared to teacher-led programs, multicomponent programs were less likely to meet SAFE criteria (65% vs. 90%) and were more likely to have implementation problems (31% vs. 22%, respectively). This creates a confound, in that multicomponent programs were less likely to contain features that were significantly associated with better results for most outcomes, and may explain why the hypothesized superiority of multicomponent programs was not confirmed.

Ruling Out Rival Hypotheses

After our primary analyses were conducted (see Table 2), we examined other possible explanations for these results. Additional analyses were conducted by collapsing across the three intervention formats and analyzing effects for the six outcome categories at post. First, we separately analyzed the impact of six methodological features (i.e., use of randomized designs, total and differential attrition, use of a reliable or valid outcome measure, and source of data: students vs. all others). We also analyzed outcomes as a function of students’ mean age, the duration of intervention (in both weeks and number of sessions), and the school’s geographical location (i.e., urban, suburban, or rural). We compared ESs for the three largest cells containing ethnicity data (Caucasian, \( k = 48 \); African American, \( k = 19 \); and Mixed, \( k = 75 \)). We also examined whether published reports yielded higher ESs than unpublished reports. Finally, we assessed if the three major intervention formats differed on any of the above variables (in addition to SAFE criteria and implementation problems) that might suggest the need for additional data analysis, but this latter procedure did not reveal any major differences across formats.

Findings. Among the 72 additional analyses we conducted (12 variables crossed with six outcomes) there were only four significant findings, a number expected based on chance. Among the methodological variables the only significant finding was that for positive social behavior: Outcome data from other sources yielded significantly higher effects than those from student self-reports. The other three significant findings were all related to the skill outcome category. Students’ mean age and program duration were significantly and negatively related to skill outcomes (\( r_s = -.27 \) and -.25), and published studies yielded significantly higher mean ESs for skills than unpublished reports. We also looked for potential differences within each of our outcome categories for ESs that were and were not adjusted for preintervention differences. The patterns of our major findings were similar (i.e., on such variables as teacher-effectiveness, use of SAFE practices, and implementation).

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

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

Interpreting Obtained ESs in Context

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

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

Table 5. Comparing Current Effect Sizes to Previous Meta-Analytic Findings for School-Age Populations

It is also possible to use Cohen’s $U_2$ index to translate the mean ES on measures of academic performance into a percentile rank for the average student in the intervention group compared to the average control student who, by definition, ranks at the 50th percentile (Institute of Education Sciences, 2008b). A mean ES of 0.27 translates into a percentile difference of 11%. In other words, the average member of the control group would demonstrate an 11 percentile gain in achievement if they had participated in an SEL program. While higher ESs in each outcome area would be even more desirable, in comparison to the results of previous research, current findings suggest that SEL programs are associated with gains across several important attitudinal, behavioral, and academic domains that are comparable to those of other interventions for youth.

Discussion

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

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

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

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

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

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

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

Developing an evidence-based intervention is an essential but insufficient condition for success; the program must also be well executed. Although many studies did not provide details on the different types of implementation problems that occurred or what conditions were in place to ensure better implementation, our findings confirm the negative influence of implementation problems on program outcomes that has been reported in meta-analyses of other youth programs (DuBois et al., 2002; Smith, Schneider, Smith, & Ananiadou, 2004; Tobler et al., 2000; Wilson, Lipsey, & Derzon, 2003). Contrary to our hypothesis, we did not find the expected additional benefit of multicomponent programs over single-component (i.e., classroom-only) programs, a finding that has been reported in other reviews of prevention and youth development interventions (Catalano et al., 2002; Greenberg et al., 2001; Tobler et al., 2000).

In the current meta-analysis, this may be due to the fact that compared to classroom-only programs, multicomponent programs were less likely to follow SAFE procedures when promoting student skills and were more likely to encounter implementation problems. It is probable that the presence of one or both of these variables reduced program impact for many multicomponent interventions. For example, many multicomponent programs involved either or both a parent and schoolwide component, and these additional elements require careful planning and integration. Others have found that more complicated and extensive programs are likely to encounter problems in implementation (Durlak & DuPre, 2008; Wilson & Lipsey, 2007; Wilson et al., 2003). It is also important to point out that few studies compared directly the effects of classroom-based programming with classroom programming plus coordinated schoolwide and parent
components (e.g., Flay, Graumlich, Segawa, Burns, & Holliday, 2004). An important priority for future research is to determine through randomized trials the extent to which additional components add value to classroom training. How much confidence can be placed in the current findings? Our general approach and analytic strategy had several strengths: the careful search for relevant published and unpublished studies, testing of a priori hypotheses, and subsequent analyses ruling out plausible alternative explanations for the findings. We also reanalyzed our initial findings to account for nested designs that could inflate Type I error rates. Furthermore, we used only school records of grades and standardized achievement test scores as measures of academic performance, not students’ self-reports, and when examining follow-up results, we required data collection to be at least 6 months postintervention. Overall, findings from the current meta-analysis point to the benefits of SEL programming. Nevertheless, current findings are not definitive. The longitudinal research of Duncan et al. (2007) presented an alternative perspective in pointing out that attention skills, but not social skills, predict achievement outcomes. They noted, however, that social-emotional competencies may predict other mediators of school success such as self-concept, school adjustment, school engagement, motivation for learning, and relationships with peers and teachers. Future research on SEL programming can be improved in several ways to shed light on if and how newly developed SEL skills in school children relate to their subsequent adjustment and academic performance.

Limitations and Future Research Directions

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

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

More attention should focus on other potential moderators of program outcomes. We evaluated the composite effects of following four recommended practices (Sequential, Active, Focused, and Explicit) relating to effective skill training because previous authors have emphasized that these factors act in combination to produce better results. However, it is possible that some practices may be more important than others depending on the nature and number of targeted skills and the developmental abilities of students. For example, younger students may need more time to acquire more complex skills. Moreover, the four practices we evaluated do not capture every aspect of effective skill development such as procedures to encourage generalization of newly learned skills and training that is developmentally and culturally appropriate (Dusenbury & Falco, 1995; Gresham, 1995). We could not examine these other features due to lack of information in study reports, but their impact on skill development merits future attention. Furthermore, it would be preferable to evaluate SAFE practices as continuous rather than dichotomous variables. That is, program staff can be compared in terms of how much they focus on skill development and the extent of their use of active learning techniques instead of viewing these practices as all-or-none phenomena. An observational system has been developed to assess the use of SAFE practices as continuous variables in youth settings (Pechman, Russell, & Birmingham, 2008). Although current results support the impact of implementation on outcomes, 43% of the studies did not monitor implementation in any way and thus were excluded from that analysis. Assessing implementation should be seen as a fundamental and necessary aspect of any future program evaluations and efforts should be undertaken to evaluate the multiple ecological factors that can hinder or promote effective delivery of new programs (Durik & Dupre, 2008; Greenhalgh et al., 2005).

Raising Healthy Children: Implications for Policy and Practice

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

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

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

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

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

Addressing these issues will increase the likelihood that more evidence-based programs will be effectively implemented and sustained in more schools, which, in turn, will support the healthy academic, social, and emotional development of more children.

References

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

  o CrossRef
  o PubMed
  o CAS
  o Web of Science® Times Cited: 66
  o CrossRef
  o CrossRef
  o CrossRef
  o CrossRef
  o CrossRef
  o CrossRef
  o CrossRef
  o PubMed
  o Web of Science® Times Cited: 18
  o CrossRef
  o PubMed
  o Web of Science® Times Cited: 4
  o CrossRef
  o Web of Science® Times Cited: 1
  o CrossRef
  o Web of Science® Times Cited: 20
  o CrossRef
  o PubMed
  o CAS
  o Web of Science® Times Cited: 382
  o CrossRef
  o PubMed
  o CAS
  o Web of Science® Times Cited: 250
  o CrossRef
  o PubMed
  o CAS
  o Web of Science® Times Cited: 136
  o CrossRef
  o PubMed
  o CAS
  o Web of Science® Times Cited: 194
  o CrossRef
  o Web of Science® Times Cited: 35
  o CrossRef
  o PubMed
  o CAS
  o Web of Science® Times Cited: 30
  o CrossRef
  o PubMed
  o CAS
  o Web of Science® Times Cited: 51
  o CrossRef
  o PubMed
  o CAS
  
  - CrossRef
  - PubMed
  - Web of Science® Times Cited: 22

  
  - CrossRef
  - Web of Science® Times Cited: 16

  
  - CrossRef

  
  - CrossRef
  - Web of Science® Times Cited: 107

  
  - CrossRef
  - PubMed
  - Web of Science® Times Cited: 14


Direct Link:
  
  - Abstract
  - PDF(213K)

  
  - CrossRef
  - PubMed
  - CAS
  - Web of Science® Times Cited: 83


Direct Link:
  
  - PubMed


Direct Link:
  
  - Abstract
  - PDF(312K)

References


• Ciechalski, J. C., & Schmidt, M. W. (1995). The effects of social skills training on students with


• Coleman, J. K. (2000). A controlled evaluation of the effects of Classroom Coping Skills training on children’s
aggressive and externalizing behaviors. Dissertation Abstracts International: Section B. Sciences and
Engineering, 61(07), 3836.

• Collaborative for Academic, Social, and Emotional Learning, (2005). Safe and sound: An educational leader’s


  o CrossRef,
  o PubMed,
  o Web of Science® Times Cited: 42

• Conduct Problems Prevention Research Group. (1999). Initial impact of the fast track prevention trial for conduct
  o CrossRef,
  o PubMed,
  o Web of Science® Times Cited: 160

school development program in Prince George’s County, Maryland: A theory-based evaluation. American
  o Web of Science® Times Cited: 43

• Cook, T. D., Murphy, R. F., & Hunt, H. D. (2000). Comer’s school development program in Chicago: A theory-
  o Web of Science® Times Cited: 34

program in the school setting: Description and evaluation. Journal of Psychology: Interdisciplinary and
  o CrossRef,
  o PubMed,
  o CAS,
  o Web of Science® Times Cited: 64

  o CrossRef,
  o PubMed,
  o CAS,
  o Web of Science® Times Cited: 66

• Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in
  o CrossRef,
  o Web of Science® Times Cited: 961

• Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of
  o CrossRef,
  o PubMed,
  o Web of Science® Times Cited: 125

• Cunningham, E. G., Brandon, C. M., & Frydenberg, E. (2002). Enhancing coping resources in early adolescence
through a school-based program teaching optimistic thinking skills. Anxiety, Stress, and Coping, 15, 369–381.
  o CrossRef,
  o Web of Science® Times Cited: 4

emotional learning: Implementation guide and toolkit. Chicago: Collaborative for Academic, Social, and Emotional
Learning.

• Diguiseppe, R., & Kassinove, H. (1976). Effects of a rational-emotive school mental health program on children’s
  o CrossRef,
  o Web of Science® Times Cited: 4

• Direct Link:
Abstract


References

Direct Link:

- [Abstract](#)
- [PDF(747K)](#)
- [References](#)


Direct Link:

- [Abstract](#)
- [PDF(877K)](#)
- [References](#)

  - [PubMed](#)


Direct Link:

- [Abstract](#)
- [PDF(681K)](#)
- [References](#)

  - [CrossRef](#)
  - [PubMed](#)
  - [CAS](#)
  - [Web of Science® Times Cited: 48](#)

  - [CrossRef](#)
  - [Web of Science® Times Cited: 4](#)

  - [CrossRef](#)
  - [PubMed](#)
  - [CAS](#)
  - [Web of Science® Times Cited: 43](#)
  - CrossRef
  - CrossRef
  - PubMed
  - CAS
  - Web of Science® Times Cited: 31
  - CrossRef
  - CrossRef
  - PubMed
  - CAS
  - Web of Science® Times Cited: 87
  - PubMed
  - CrossRef
  - PubMed
  - CAS
  - CrossRef
  - PubMed
  - CAS
  - Web of Science® Times Cited: 62
  - CrossRef
  - PubMed
  - CAS
  - Web of Science® Times Cited: 78
  - CrossRef
  - PubMed
  - Web of Science® Times Cited: 43
  - CrossRef
  - PubMed
  - Web of Science® Times Cited: 17
  o CrossRef
  o PubMed
  o CAS
  o Web of Science® Times Cited: 32
  o CrossRef
  o PubMed
  o Web of Science® Times Cited: 32
  o CrossRef
  o Web of Science® Times Cited: 17
  o PubMed
  o CAS
  o Web of Science® Times Cited: 28
  o CrossRef
  o CrossRef
  o Web of Science® Times Cited: 32
  o CrossRef
  o PubMed
  o CAS
  o Web of Science® Times Cited: 19
  o Abstract
  o PDF (1386K)
  o References
  o CrossRef
  o Web of Science® Times Cited: 18
  o CrossRef
  o Web of Science® Times Cited: 66
  o Web of Science® Times Cited: 68

Direct Link:

- CrossRef
- PubMed

References

  - CrossRef
  - PubMed

  - PubMed

  - CrossRef
  - Web of Science® Times Cited: 172

  - CrossRef
  - PubMed
  - Web of Science® Times Cited: 140

  - CrossRef


  - CrossRef
  - PubMed
  - Web of Science® Times Cited: 14

  - CrossRef
  - PubMed
  - CAS
  - Web of Science® Times Cited: 164


Direct Link:

- Abstract
- PDF(97K)

References

  - CrossRef
  - PubMed
  - Web of Science® Times Cited: 22


• **Heinemann, G. H.** (1990). The effects of the Lions-Quest “Skills for Adolescence” program on self-esteem development and academic achievement at the middle school level. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 51*(06), 1890A.


Direct Link:
- **Abstract**
- **Full Article (HTML)**
- **PDF(99K)**
References


Abstract


References


- Web of Science® Times Cited: 32
- CrossRef
- PubMed
- CAS


- CrossRef
- PubMed
- CAS


- Web of Science® Times Cited: 1


- CrossRef
- PubMed


- CrossRef
- Web of Science® Times Cited: 37


- Abstract
- PDF(138K)
- References


- CrossRef
- PubMed
- Web of Science® Times Cited: 49


- CrossRef
- PubMed
- Web of Science® Times Cited: 73


Direct Link:

- Abstract
- PDF(646K)
- References
  - CrossRef
  - PubMed
  - CAS
  - Web of Science® Times Cited: 18

  - CrossRef
  - PubMed
  - Web of Science® Times Cited: 35

  - PubMed


Direct Link:
  - PDF (969K)
  - References

  - CrossRef


Direct Link:
  - PDF (2240K)
  - References

  - Web of Science® Times Cited: 6

  - CrossRef
  - PubMed
  - CAS
  - Web of Science® Times Cited: 23

  - CrossRef
  - PubMed
  - CAS
  - Web of Science® Times Cited: 35


  - CrossRef
  - PubMed
  - Web of Science® Times Cited: 28

References


Abstract

  Direct Link:
  o Abstract
  o PDF(519K)
  o References
  o PubMed
  o CAS
  o Web of Science® Times Cited: 5
  o CrossRef
  o PubMed
  o CAS
  o Web of Science® Times Cited: 145
  o Web of Science® Times Cited: 5
  o CrossRef
  o PubMed
  o CAS
  o Web of Science® Times Cited: 49
  Direct Link:
  o Abstract
  o PDF(556K)
  o References
  o Web of Science® Times Cited: 7
  Direct Link:
  o Abstract
  o PDF(528K)
  o References
  o CrossRef
  o Web of Science® Times Cited: 24
  Direct Link:
  o Abstract
  o PDF(69K)
  o References
  o CrossRef
  o PubMed


Direct Link:
  o Abstract
  o PDF(219K)
  o Web of Science® Times Cited: 13
  o CrossRef
  o Web of Science® Times Cited: 21
  o Web of Science® Times Cited: 4
  o CrossRef
  o PubMed
  o CAS
  o Web of Science® Times Cited: 3
  o CrossRef
  o Web of Science® Times Cited: 6
  o CrossRef
  o CrossRef
  o PubMed
  o CAS
  o Web of Science® Times Cited: 109
  o CrossRef
  o PubMed
  o CrossRef
  o PubMed
  o CAS
  o Web of Science® Times Cited: 42
  o Web of Science® Times Cited: 18
  o CrossRef
Direct Link:
Abstract


References

- **PubMed.**
- **CAS.**
- **Web of Science® Times Cited: 6**
- **CrossRef.**
- **PubMed.**
- **Web of Science® Times Cited: 74**
- **CrossRef.**
- **Web of Science® Times Cited: 213**
- **CrossRef**
- **CrossRef.**
- **Web of Science® Times Cited: 55**
- **CrossRef.**
- **Web of Science® Times Cited: 68**
- **CrossRef.**
- **PubMed.**
- **Web of Science® Times Cited: 81**
- **CrossRef.**
- **PubMed.**
- **Web of Science® Times Cited: 74**
Direct Link:
- **Abstract**
- **PDF(277K)**
- References
- **CrossRef.**
- **Web of Science® Times Cited: 87**


THE ECONOMIC VALUE OF SOCIAL AND EMOTIONAL LEARNING

February 2015 (Revised)

Clive Belfield
Brooks Bowden
Alli Klapp
Henry Levin
Robert Shand
Sabine Zander

Center for Benefit-Cost Studies in Education
Teachers College, Columbia University
www.cbcse.org

Funding for this research was provided by the NoVo Foundation (novofoundation.org).
The Economic Value of Social and Emotional Learning
Clive Belfield, Brooks Bowden, Alli Klapp, Henry Levin, Robert Shand and Sabine Zander

Center for Benefit-Cost Studies in Education
Teachers College, Columbia University
www.cbcse.org

February 2015 (REVISED VERSION)

Funding for this research was provided by the NoVo Foundation (novofoundation.org).

Acknowledgements: The authors greatly appreciate comments and support from Steve Aos, John Bridgeland, Mark Greenberg, Damon Jones, Stephanie Lee, Annie Pennucci, and Roger Weissberg.
Benefit-cost analysis is a tool for evaluating the economic profitability of an investment. It has been used in education since the 1960’s to determine the rate of return on both individual and social investments in education. Essentially, benefit-cost analysis compares the monetary cost of an investment with the monetary value of its outcomes. For example, by reducing high school dropouts there are costs to the student in foregone income by staying out of the labor market. But there are also gains to the student in terms of higher income, better health, and lower likelihood of involvement in the criminal justice system, all which can be measured, in terms of a monetary return on investment. The taxpayer also makes an investment in education through paying a considerable portion of its direct cost and gets a return through higher tax revenues and lower costs of public services for health, public assistance, and criminal justice. And, society obtains returns by using its resources in its most productive ways, at least partially reflected in economic returns.

It is not necessarily the case that the benefits of an educational investment always justify the costs. For some interventions, the benefits will exceed the costs. But in others the investment is not found to be compensated by its returns: even if an intervention is effective, it may be too expensive to implement. Thus, the purpose of a benefit-cost (BC) study is to provide an accurate measure of costs and benefits to determine not only if the benefits are greater than the costs of the investment, but by how much.

In the educational setting, we seek investments that have the highest return to the taxpayer and to society. In the past, BC studies have been limited largely to increases in educational attainment and to improvements in cognitive test scores. But it is now becoming widely recognized that social and emotional learning in schools can be as important as or even more important than cognitive gains in explaining important developmental and life outcomes (Durlak et al. 2011; Heckman and Kautz 2012; Levin 2012). Social and emotional skills are less commonly considered in educational evaluations, in part because they are more challenging to measure than attainment and test scores. As such skills have gained prominence, it is important to integrate them into BC studies for consideration in educational policy and decision-making.

Because of the long history of the Center for Benefit-Cost Studies in Education (CBCSE) at Teachers College and its predecessor organizations in publishing BC studies in education (e.g. Levin 1975, Levin and McEwan 2001, and other publications at www.cbcse.org), it was requested by the NoVo Foundation and the Collaborative for Academic, Social, and Emotional Learning (CASEL) to apply BC analyses to studies of social and emotional learning. The purpose was to both demonstrate the BC method in this domain as well as to gain early perspectives on the potential economic returns to investments in social and emotional learning.

In this Report, we review the available evidence on the economic value of Social and Emotional Learning (SEL). We utilize a formal method to perform economic valuations with respect to changes
in Social and Emotional (SE) skills to guide future evaluations of reforms that target SEL. Our main contribution is to demonstrate BC analysis using recent impact evaluations of six prominent SE intervention:

- 4Rs;
- Positive Action;
- Life Skills Training;
- Second Step;
- Responsive Classroom; and
- Social and Emotional Training (Sweden).

Table S1 provides the descriptions of the six interventions.

These interventions were chosen because they are prominent in the literature and provide diversity in terms of their goals, measures of outcomes, and student populations. For each intervention we have constructed tables of ingredients and their costs; alongside, we have created benefit maps to summarize the possible benefits each intervention might confer and calculated the monetary value of the portion of benefits that could be identified and quantified based on the results reported in the impact evaluation. We then computed appropriate economic metrics – benefit-cost ratios and net present values – and performed sensitivity testing to see if the results are robust to alternative specifications.

Our central question is whether a range of different SEL interventions, both individually and in the aggregate, show benefits that exceed their costs. We recognize the fact that we have captured only a portion of their benefits because not all of their effects are fully measured and can be readily converted into monetary measures of their benefits. Further, the portion of benefits that we have been
able to capture may differ considerably among the different interventions' depending upon the goals of each and measurability of the associated benefits. We caution that it is inappropriate to compare individual benefit-cost ratios among the six interventions in the absence of a more complete accounting of benefits for each. Thus, we focus primarily on the basic question of whether the available measures of benefits are equal to or exceed costs for each and summarize our results with an overall average benefit-cost ratio among the six interventions.

The most important empirical finding is that each of the six interventions for improving SEL shows measurable benefits that exceed its costs, often by considerable amounts. There is a positive return on investments for all of these educational reforms on social and emotional learning. And the aggregate result also shows considerable benefits relative to costs, with an average benefit-cost ratio of about 11 to 1 among the six interventions. This means that, on average, for every dollar invested equally across the six SEL interventions, there is a return of eleven dollars, a substantial economic return. These findings are robust to the imposition of different assumptions on the sources and construction of benefits and costs, and a full accounting for benefits, as shown in the benefit maps, would provide an even larger return.

However, we emphasize that in addition to the benefit-cost test that we have imposed on these interventions, an important contribution of this report is its demonstration of methods and applications to estimating benefits and costs of SEL interventions. Only if a formal method is applied in a transparent way will economic evaluations of SEL interventions help improve resource allocation within the school setting.

This methodological demonstration can advance understanding of SEL in the future. Because our calculations have not captured all of the benefits of each of the interventions and the proportions that we have captured probably vary considerably among them, we caution the reader that it is inappropriate to compare directly the results for specific interventions relative to the others. The important findings are that each of the SEL approaches shows benefits that exceed costs and that the average return is very high, but differences in the quality and availability of comparable data preclude precise comparisons among the interventions for the following reasons:

- **Evaluation Quality**—We drew upon the best evaluations for each intervention that were available. However, even among relatively high quality evaluations there was variation. Had the evaluation designs and measurements been of consistent quality, there might have been some differences in results. Both data quality and research methods account for some potential differences in our results.

- **Cost Estimation**—Our cost method is the well-established ingredients method, which requires identification of all of the resources that were required to produce the results that were found in the evaluations. Since all evaluations had been done in the past, it was necessary to reconstruct the interventions as they had been implemented at the time of their reported effectiveness results. Precise details of the interventions were rarely available, and attempts to contact informed observers were not always successful. Even when successful, relying on retrospection of details can often provide incomplete or misleading results. To as great an extent as possible we have attempted to use multiple sources to construct the cost estimates, but differences in the information base attainable for each of the interventions varied considerably in comprehensiveness and quality.

- **Multiple Benefits**—Most SEL interventions have multiple goals and benefits, and this contrasts with interventions to improve cognitive test results in a particular subject. For example, an attempt to reduce aggression may also improve impulse control and later reduce juvenile crime and/or may
raise academic achievement. We have tried to capture some of this heterogeneity in our benefit maps, but even these are dependent upon an incomplete knowledge-base. And as the maps reveal, only a portion of the possible benefits are measured in the specific evaluation studies. Consequently, our measures of benefits are based upon a limited set of dimensions, a source of downward bias in our benefits measure. The actual benefits may be considerably higher if we were able to identify all effects and convert them into monetary benefits.

- **Benefits Measures**—There is a major challenge in taking outcomes such as reductions in substance abuse or aggression or other improvements in behavior and attitudes and converting them into monetary values to society. We have used shadow prices, which reflect the amount that society is willing to pay for such improvements. But shadow prices depend upon specific assumptions, constructs, and data availability, and there is likely to be some underlying variance, largely unmeasured, in the magnitude of potential shadow price estimates. We believe that the shadow prices used to estimate benefits in this report are both plausible and defensible, but they are still subject to variability.

Taken in conjunction, the two contributions of this Report suggest a promising agenda for future research on the economics of social and emotional learning. From one perspective, it is feasible to apply benefit-cost analysis to SEL interventions, and these interventions do offer high economic returns as educational investments. Overall, SEL interventions are likely to pass a benefit-cost test. From another perspective, there is considerable additional research to be performed to establish the full extent and magnitude of the benefits of SEL. The full economic value of SEL is not yet established. Our hope is, therefore, that this investment stimulates the generation of a refined knowledge base and a greater focus on the development of benefit-cost applications to SEL initiatives.
Summary
1. Introduction
2. Evidence on the Economics of SEL
   2.1 Program-Based Evidence
   2.2 Earnings and SE Skills
   2.3 Educational Achievement and SE Skills
3. Foundations for Economic Evaluation of SEL
   3.1 Economic Framework
   3.2 Ingredients Method for Costs
   3.3 Shadow Price Methods for Benefits
4. SEL Interventions
   4.1 4Rs
   4.2 Positive Action
   4.3 Life Skills Training
   4.4 Second Step
   4.5 Responsive Classroom
   4.6 Social and Emotional Training
5. Developing Methods and Evidence for Economic Evaluations of SEL
References
Appendices
Appendix I Benefits Maps
Appendix II
1. INTRODUCTION

Social and emotional skills are increasingly being recognized as important for child development (Weissberg et al., 2003; Zins et al., 2004). These skills, which include competencies such as self-awareness, self-management, social awareness, relationship skills, and responsible decision-making, may help students progress further in their education and may also enhance personal, economic and social well-being in youth and adulthood. In recent reviews of the evidence on interventions to develop social and emotional skills, Durlak et al. (2011, Table 2) and Sklad et al. (2012) identify durable and substantively important gains in social and emotional skills, attitudes, positive social behavior, conduct problems, emotional distress and academic performance. These gains should lead to substantial increases in personal and social well-being.

According to Durlak et al. (2011), social and emotional learning combines youth development with the promotion of particular competencies, with the aim of enabling students to respond appropriately to environmental demands and fully take advantage of opportunities. Ultimately, social and emotional competencies encourage a shift to an internal locus of control, allowing individuals’ choices and actions to better accord with their own values. These competencies relate to “soft skills” and personality traits that, according to Heckman and Kautz (2012), predict success in school, the labor market, and in life. Notably, social and emotional competencies do not just raise academic achievement and educational attainment. They also foster personal satisfaction and growth, help individuals become better citizens, and reduce risky behaviors like violence and drug use (Durlak, Weissberg and Pachan, 2010; Collaborative for Academic, Social, and Emotional Learning, 2013).

However, although the positive development of social and emotional skills through education has been established, the economic value of gains in these skills has received little attention. But, within a resource-constrained education system, instructional activities in social and emotional skills must compete with instruction directed toward test score improvements and increases in educational attainment. To date, most education research has focused on these last two objectives, with the presumption that this will have important economic consequences. For attainment, this focus makes sense: the evidence on the high economic value of graduating from high school or completing college is substantial. For achievement, however, the justification is much less certain: the impact of cognitive gains on economic performance is extremely modest; and most interventions fail to generate long-term boosts in cognition (Levin, 2012). Moreover, whatever influences are driving the high returns to attainment, they are only partially mediated through improvements in cognitive functioning (Heckman and Kautz, 2012). A suggestive and plausible alternative is that much of the effect of education on economic outcomes and personal well-being is based on non-cognitive or social and emotional development such as effort, motivation, curiosity, empathy, caring, and the many other dimensions that can be encompassed by social and emotional skills. These skills may drive attainment, which then becomes the focus for educational investments. On balance, it is likely that both social and emotional learning and cognitive learning are efficient investments. But the former domain has received very little attention from economists, and the policy debate instead emphasizes cognitive gains and test score accountability. In this environment, it is critical to identify and enumerate the value of social and emotional skills.

In this Report, we demonstrate an approach to conducting economic evaluations of six school-based social and emotional learning interventions and place these evaluations in the broader context of the economics of social and emotional learning. These six interventions are selected because each
has been evaluated using a generally, acceptable, research method in terms of impacts on a range of social and emotional (SE hereinafter) skills. We estimate each intervention’s costs based on the ingredients employed during the implementation previously evaluated. We utilize the effects estimated in the evaluations to estimate economic benefits of the interventions to society. We then calculate the benefit-cost ratios and net present values to determine if the benefits generated by each program outweigh the costs of implementation. The intention here is not to rank these interventions in terms of their efficiency for reasons of differential completeness of the underlying studies in identifying effects and measurement challenges in identifying benefits. Instead, it is to provide a demonstration of how benefit-cost analysis of social and emotional interventions should be undertaken and to see if measured benefits exceed costs, understanding that a full assessment of benefits of each intervention would yield even higher benefits. Thus, we pay considerable attention to methodological challenges in estimating costs and calculating benefits; in turn, these challenges highlight broader research issues in evaluating social and emotional interventions.

Our Report is structured as follows. We begin with a review the evidence on the development of SE skills and in particular how these skills influence future earnings and are mediated through changes in education levels. Next, we set out a framework that includes both the economic principles and the methodological approaches for estimating the value of SE skills. We then apply this framework to six SE interventions to calculate their costs, benefits, and net benefits. Finally, in light of these six applications, we illustrate important conceptual, empirical, and methodological issues for research on the economic value of SEL. Overall, by providing the foundations for economic evaluations of social and emotional skills, this research should enhance their presence in future research and policy discussions.
2. EVIDENCE ON THE ECONOMICS OF SEL

2.1 Program-Based Evidence

There is substantial evidence on positive impacts of SEL interventions (SCDRC, 2010; Durlak et al., 2011; and Sklad et al., 2012). These impacts include large gains in SE skills, attitudes, positive social behavior, and academic performance, as well as reduced conduct problems of student disruption and emotional distress. These gains, measured across hundreds of interventions, are substantively large (with effect size gains of 0.2-0.6); and, when based on teacher reports, are consistently statistically significant (Durlak et al., 2011, Table 2). In a subsequent review, Sklad et al. (2012) emphasize gains in social-emotional skills and positive self-image. Concomitant with these gains were other effects in terms of social behavior, substance abuse, mental health disorders, and achievement.

SEL interventions vary significantly in application and mission: some are integrated into regular classroom instruction, others are directed at a specific adverse behavior (e.g. substance abuse), and others are intended to have a general effect (e.g. on school climate) rather than individual students. SEL interventions can be targeted to specific grades or be applied across the full span of grades. Variations in the practices for each intervention with respect to lessons, classroom strategies, and professional development are given in detail in SCDRC (2010, Table 1.1, Panels 1-3) and Sklad et al. (2012, Tables 3 and 4). Hence, although we refer to SEL interventions as a group, we recognize there is considerable diversity.

Some benefit-cost analyses of SEL interventions have been performed.1 Applied to the context in Washington state, Lee et al. (2012) have calculated costs, benefits, and net benefits for three well-known SEL interventions: Promoting Alternative Thinking Strategies program (PATHS); the Seattle Social Development Project (SSDP); and Life Skills Training (LST).2 For the SSDP, the participant costs are $3,030 and the benefits are $5,800; the net benefits are strongly positive at $2,770. For LST, the costs are $30 and the benefits are $1,290; the net benefits are therefore $1,260. At the national level, Jones et al. (2008) estimate net benefits of $810 per student in LST. Looking at how SEL interventions ameliorate substance abuse by youth, Miller and Hendrie (2008) perform benefit-cost analysis for three interventions. For LST, they estimate costs of $290, benefits of $5,960, and therefore net benefits of $5,670. For SSDP, they estimate costs of $3,200, benefits of $19,000, and therefore net benefits of $15,800. Finally, for the Social Competence Program, Miller and Hendrie (2008) estimate costs of $350, benefits of $2,500, and therefore net benefits of $2,150. Given the approaches used, these benefit-cost analyses show that SEL interventions can yield positive, sometimes substantial, economic returns. However, since the studies were done independently, they use different approaches to both cost and benefit measurement. Therefore, they are not directly comparable. Across the broader literature on youth behavior, there are many benefit-cost analyses of programs to combat delinquency.3 Many of these programs are intended to modify some SE skills (such as student conduct, educational progress, and externalizing behaviors). But their main focus is on a narrow class of behaviors, such as teenage pregnancy or juvenile crime; few are intended to affect a broad array of outcomes simultaneously.

---

1 Throughout, we report all money values in 2013 dollars and, where appropriate, express amounts in present values at the start of each SEL intervention. To avoid spurious precision, amounts are rounded to the nearest $10 or $100 where appropriate.
2 See http://www.wsipp.wa.gov/BenefitCost. Evidence on PATHS is currently under review.
3 See Weimer and Vining (2009) and the compendia of evidence at: www.colorado.edu/cspv/blueprints/; youthinfo.gov; and wsipp.wa.gov.
Nonetheless, these programs are typically found to have benefits that exceed their costs. Miller and Hendrie (2008, Table 12) estimate net benefits for 17 general substance abuse programs and find almost all have positive net benefits. In their review for Washington State of programs for children and adolescents, Lee et al. (2012) find positive net benefits for seven.

Although these analyses contribute to the debate on the economic value of SEL, the literature and evidence base are limited. More importantly, significant empirical and methodological challenges remain. First, few studies report costs in detail. Even fewer use the ingredients method, relying instead on more casual information such as budgetary data or estimates by the program developers. Second, it is not clear if these programs can be compared either to each other or to the broader class of delinquency prevention interventions. There are significant programmatic differences with respect to the students targeted, the size of the program, and the outcomes intended for improvement (SCDRC, 2010, Table 1.1, Panel 6). There are also significant differences in how the programs are evaluated. These differences include: the number of impacts that are converted into monetary benefits; the time horizon for analysis (with some studies including only immediate benefits and others only long-term benefits); and the method by which benefits are calculated (i.e. how impacts are translated into dollar amounts). Unless the method is consistently applied, comparing results across existing evaluations is highly problematic.

The economic evidence on the value of SE skills therefore needs to be expanded to include more interventions and particularly those that affect SE skills directly (rather than youth conduct). The research also needs to be undertaken in a standardized way, such that findings from separate benefit-cost analyses are comparable and so can be generalized. We discuss these issues in more detail in our conclusion.

2.2 Earnings and SE Skills

One approach to establishing the value of SE skills is to look at how they directly influence labor market outcomes. There is now an expanding literature on the association between SE skills and subsequent earnings (Murnane et al., 2001; Waddell, 2006; Drago, 2011). This evidence indicates strong positive impacts on earnings when SE skills are enhanced.

However, a critical aspect of this research is how SE skills are defined and measured. These skills can be defined in terms of psychological traits, behaviors, attitudes towards others, and performance on specific tasks. In turn, these constructs can be measured in different contexts (e.g. school versus home) and using different instruments. Yet, labor market research relies almost exclusively on the Rosenberg scale of self-esteem and Rotter scale for locus of control. These scales are rarely used in evaluations of particular interventions. Thus, it is not possible to directly translate the impact of an intervention into gains in earnings.

In addition, there are methodological challenges to identifying earnings gains from social and emotional (SE) skills. One challenge is that the construct chosen to measure SE skills is typically the only one representing an individual’s behavioral traits; it may be that other traits – correlated with SE

---

4 According to Durlak et al. (2011, 413), the average follow-up period for SEL interventions is less than two years. Thus, in most cases it is not possible to see whether there have been long-run impacts. Yet, it seems likely that some SEL effects will persist, even if the immediate effects are stronger than those over a longer-term (Sklad et al., 2012). It is possible to extrapolate from immediate impacts to future ones, although this extrapolation adds imprecision.
skills – are driving the association. Another challenge is that education itself may be positively correlated with SE measures; and education is of course a significant determinant of earnings. Both these challenges may offset any underlying association. De Araujo and Lagos (2013) examine how earnings are influenced by self-esteem (their construct for SE skill); they find that self-esteem has no influence after controlling for locus of control attributes and for education. Nevertheless, if self-esteem affects educational attainment and thence earnings, this still represents an economic benefit from SE skills. A final challenge arises when these SEL constructs are measured simultaneously with earnings: it seems plausible that someone with high earnings might report (or have acquired) high SE skills. On the last of these, de Araujo and Lagos (2013) find that high self-esteem is indeed jointly determined with wages.

Notwithstanding these challenges, research evidence consistently finds strong associations between SE skills – broadly defined – and earnings. Using NLSY79, de Araujo and Lagos (2013, p.1987) estimate that a “one standard deviation increase in [self-reported] self-esteem leads to a 30.46% increase in real wages”. This increase is mostly mediated through attainment, which is estimated to be 1.5 years greater for those with higher self-esteem. If we assume baseline returns to attainment at 10% per year, the estimated gain in earnings from a one standard deviation increase in self-esteem is therefore 15%. This is still a very large gain in earnings. However, Drago (2011), also using NLSY79, finds smaller estimates of self-esteem (at 4%). Using NELS88, Segal (2013, p.767) reports a one standard deviation increase in “[teacher-reported] misbehavior is associated with a 4% decrease in earnings” at age 28. Also using NELS88, Eren and Ozbeklik (2013) employ the Rosenberg and Rotter scales in 10th grade and estimate that a one-standard deviation increase in non-cognitive ability is associated with an increase of 9% in earnings.5

Although promising, a precise association between SE skills and earnings remains to be determined. The economic value of these impacts of enhanced SE skills can only be generalized under two restrictive assumptions. It is necessary to assume that these constructs (self-esteem, misbehavior, self-control, etc.) accurately and fully reflect SE skills. As well, it is necessary to assume that SE skill differences are stable over childhood (such that SE skills in one grade reflect SE skills throughout the school years). These assumptions – both of which are highly debatable – represent important areas for future research.

Yet, given these two assumptions, it is possible to approximate the present value of lifetime earnings gain from a one standard deviation increase in SE skills. (We use one standard deviation for illustrative purposes; most interventions cannot yield effects this large). Based on current lifetime earnings profiles from the American Community Survey and the Current Population Survey, the average present value lifetime earnings of a child who is currently in 3rd grade would be $575,000.6 These amounts represent how much a lifetime's earnings are worth at 3rd grade (adjusting for inflation). Given the above evidence, if SEL goes up by one standard deviation, earnings go up by 4-15% (net of additional schooling). Using the lower bound of these numbers, the average expected earnings gain from a one standard deviation increase in SE skill in 3rd grade is $46,000. A conservative estimate – using the weakest correlation between SE skills and earnings – would yield earnings gains of $23,000. In other words, if an intervention raised a child's SE skills at that age, this would be worth $23,000 in terms of gains in future productivity alone. Although not precise and subject to the many caveats listed above, these figures indicate very large long-term labor market effects from enhanced SE skills.

5 Results from Murnane et al. (2001) and Waddell (2006) are not easily translated into effect size gains.
6 Details of this calculation are available from the authors. Under these two assumptions, the present value amount can be calculated for any year group.
2.3 Educational Achievement and SE Skills

An alternative approach to estimating the economic value of SE skills is to examine how they affect education levels. From this association it should then be possible to draw on the vast amount of research on the economic returns to education to establish the benefits of SE skills. Overall, research evidence typically shows a positive association between SE skills and education. In their review of 35 SEL interventions, Durlak et al. (2011, Table 2) report an average effect size gain in achievement of 0.27.

From an economic perspective the association between SE skills and achievement is complex. Where SE skills enhance achievement, it should be possible to value SE skills by valuing the labor market gains from higher achievement. However, this approach is likely to undervalue SE skills as these almost certainly have much more diffuse and long-term behavioral and attitudinal implications than achievement does. Academic achievement gains often fade-out within a short period (see Kinsler, 2012; Jacob et al., 2008). Moreover, for some groups academic achievement per se may not be especially important for future lifetime outcomes and may even be inversely related to SE skills (as established in studies of high school dropouts who pass the GED, see Heckman and LaFontaine, 2006). Recent research using longitudinal data by Heckman and Kautz (2012) describes a weak association between achievement and future life outcomes; and Castex and Dechter (2013) find that cognitive ability has actually declined as a determinant of wages over the last two decades. Heckman and Kautz (2012) also argue that SE skills and achievement as measured by standardized test scores are strongly confounded: because SE skills are expressible as behavior on tasks and taking an achievement test is a task, then higher achievement scores must also reflect higher SE skills. From an economic perspective, distinguishing the two constructs is problematic and may lead to double-counting in benefit-cost analysis because of their overlap. Finally, it might be noted that a focus on the cognitive gains of SEL programs serves to reinforce the idea that schools teach only ‘knowledge’ and downplays the idea that schools ought to develop broader social and emotional skills.

A number of studies have – despite these caveats – evaluated SE programs in terms of how they increase achievement, and these achievement gains can be translated into present value money benefits in terms of higher earnings. (As with the earnings analyses referred to above, an implicit assumption is that achievement gains are stable and persist over grade levels). To estimate the economic gains from achievement, we adapt estimates generated by Belfield and Levin (2009, Table 4). Specifically, we adjust their estimates of achievement gains – mediated through changes in attainment – for inflation and population weights. These adjustments yield a present value gain in earnings from a one standard deviation increase in 3rd [8th] grade math test scores of $34,300 [$40,700]. Applied to the specific results reported by Durlak et al. (2011), a one standard deviation increase in SE skills in 3rd [8th] grade is therefore valued at $9,000 [$11,000]. Although still substantial, these education-mediated values are considerably below the direct estimates reported above.

Thus, the association between SE skills and education should be evaluated more broadly than through individual achievement gains. Although Zins et al. (2007) make a strong case for evaluating SE skills in terms of academic outcomes, these outcomes should include school behaviors and school attitudes, as well as test scores (Zins et al., 2007, Figure 1.4). One direct way that has clear economic

---

7 To avoid the methodological challenges noted above, Belfield and Levin (2009) express achievement gains in terms of attainment gains (using the association between test scores and high school graduation).
consequences is to look at rates of grade retention and placement in special education. So far, this approach – even as it only refers to substantive behaviors and not attitudes – has not been extensively investigated. Moreover, it seems unlikely that the benefits of SE are confined solely to the individual student. The benefits of SE skills should be looked at using general measures at both classroom-level (such as peer effects, teacher efficacy, or teacher retention) and school-level (such as school quality or school climate). That is, in an economic sense there are externalities such that a healthier emotional and social climate induced directly by SEL interventions at the level of individual students should improve classroom harmony and school functioning. Unfortunately, to our knowledge there is no straightforward way to estimate the economic value of general school quality/climate differences. It seems plausible that in the absence of SEL, schools would incur greater costs in terms of heightened security, student support, or facilities upkeep. Related evidence on ‘compensating wage differentials’ (in this case, the need to pay teachers more to work in more disruptive school environments) has not found large costs when students exhibit worse behavior (Goldhaber et al., 2010). At this time it is therefore possible to calculate the economic value of SE skills – as mediated through education – only in terms of individual test score gains or more productive social behaviors and not in terms of other school-wide or student-level effects.

---

8 Again, this task is made more difficult by the idiosyncratic instruments used by evaluators to measure school climate.
3. FOUNDATIONS FOR ECONOMIC EVALUATION OF SEL

3.1 Economic Framework

A framework for economic evaluation of SEL interventions – linking costs and benefits together in a standardized approach – is given in Figure 1.

To begin, the costs of each intervention must be estimated. These costs should be based on the opportunity cost of the resources required (Levin, 1975; Nas, 1996, p.61). They should be expressed in present values (discounted) from the start-date of the intervention to account for the fact that money expenditures that happen earlier as reflected in opportunity cost are given greater weight by society (see McEwan, 2002, p.39).

Next, the benefits of the intervention must be estimated, ideally based on causal impact evaluations. These too should be expressed as present values at the start date of the intervention: present investments in SEL engender future benefits and the further these benefits are in the future, the less valuable they are. By convention, a societal perspective is adopted whereby all benefits are calculated regardless of who reaps them. These benefits can be separated into three components. First, there are immediate benefits $B_{Di}$ – those occurring as the intervention is being delivered. Second, there are post-intervention benefits during youth $B_{yi}$ – those benefits that occur after the intervention has been delivered but while the participants are still in school. Third, there are post-intervention benefits during adulthood $B_{Ai}$ – those benefits that occur after the participants have left school. There may be many separate benefits in each time period and each benefit must be modeled over time. For the immediate benefits, the issue is how rapidly the benefits accrue once the intervention has begun (specified as the ‘ratchet function’ $r_i$). Some interventions may work immediately (perhaps through control of behaviors); others may work more slowly (perhaps through changing attitudes). Holding all other factors constant, immediate benefits are preferred to later ones. For the post-intervention benefits, the critical issue is their durability (‘fade-out function’ $f_i$). Some interventions may yield benefits only during implementation; others may change behaviors beyond the time frame of intervention. Critically, all benefits from an intervention should be counted: otherwise – tautologically – analyses that include more benefits will appear more beneficial.

Costs and benefits need to be reported in an equivalent way. For these analyses, all dollar amounts are reported as present values of future benefits from the year in which the intervention began with a discount.
rate of 3.5% applied to future costs and benefits. Dollar values are adjusted for inflation and reported in 2013 prices. Unless otherwise stated, the general ratchet function is assumed to be zero (i.e. the impacts of the program only occur in the year in which they are measured). Unless otherwise stated, the fade-out function is zero: the benefits are assumed to persist through school and adulthood. (Although this may seem to be a strict assumption, it is unlikely that interventions are delivered under the assumption that they will only have temporary effects. In our sensitivity testing, we vary the rate of fade-out).

Two efficiency metrics can be derived from comparing costs and benefits: the benefit–cost ratio \((= B/C)\) and the net benefits \((= B-C)\).\(^9\) Interventions with benefit-cost ratios that exceed one or for which benefits exceed the costs are considered efficient. Interventions with the highest net benefits or benefit-cost ratios are most preferable from an efficiency standpoint. Given that SEL interventions are often at substantially different scales of implementation, the net benefit measure may be more informative. The net benefit measure indicates the total amount of resource saved: interventions that generate the largest total dollar saving are preferred (even if they have a lower benefit-cost ratio). Calculation of these metrics is subject to sensitivity testing. In this analysis, a general form of sensitivity analysis is performed: salient benefits and costs that cannot be monetized are mapped, along with analysis as to how their inclusion would influence the results.

It is important to specify this framework in detail because interventions can differ in important respects. Most obviously, they can differ in how much resource is allocated to them: how much they cost and who bears that cost. Next, they can differ in the number and types of separate domains of behavior they impact. Those interventions that reduce juvenile crime and substance abuse will be more beneficial than those that impact in just one of these domains. Also, interventions can differ with respect to the durability of their impacts. Those that yield short-term gains in academic achievement are unlikely to be as beneficial as those that ensure students graduate from high school, for example. In conjunction, there are a series of context assumptions (year of evaluation, inflation adjustments) that can create variation in analyses. These differences have to be ascertained when comparing interventions.

Application of this framework – consistent measurement of costs and benefits, reporting of standardized metrics, and consideration of differences – should yield meaningful estimates of the economic value of interventions to boost SE skills.

Moreover, this framework gains even greater salience in light of current practice in SEL evaluations. For most interventions there are important uncertainties and knowledge gaps. Most evaluations give a very incomplete sense of how valuable – and hence how attractive to policymakers – a particular intervention is. This economic framework serves to illustrate these uncertainties and gaps. As such cost-benefit analysis yields information on what we do know and illustrates the possible importance of what we do not know.

### 3.2 Ingredients Method for Costs

The ingredients method is a cost-accounting approach that is compatible with the economic concept of opportunity cost and has been used successfully to undertake cost and cost-effectiveness studies in education (Levin and McEwan, 2001; IOM/NRC, 2014). With this method all the ingredients or

\(^9\) A third metric is the internal rate of return, defined as the discount rate that makes the present value of benefits equal to the present value of costs. This metric conveys less information than the other two metrics but may be useful as a shorthand way to compare programs of similar scale and duration.
inputs used to implement an intervention – net of the costs of the alternative or ‘business as usual’ program – are identified and specified. Inputs should be counted as incremental beyond what is usually required in the absence of the intervention. All inputs should be included and evaluated for their costs, regardless of who pays for them (with their financing source identified).

Program ingredients are grouped into personnel, materials/equipment, facilities and other inputs. Ideally, information on ingredients is collected from semi-structured instruments administered to key personnel responsible for implementing the intervention. Each ingredient is then priced out using prices from independent sources based on actual market outcomes (or shadow prices if market prices do not reflect opportunity costs or are unavailable). The cost of the intervention is then reported as the total sum of all ingredients multiplied by their unit prices. Also, some interventions induce additional resource use if the intervention succeeds in improving earlier educational outcomes (e.g., if more students are induced to attend college). When monetary benefits are calculated, additional costs engendered by success are typically counted as ‘negative benefits’ and put on the benefits side of the ledger.

The ingredients costing method is strongly preferred over reliance on budgets (and very strongly preferred over reliance on statements from program deliverers). Agency budgets are rarely comprehensive, covering all relevant ingredients, and are almost never itemized in a way that clearly identifies how much is spent on a given intervention. Using budgets it is very difficult to disentangle the incremental costs of an intervention from the regular school operations. As well, budget statements reflect local prices and not what an intervention would cost if another agency decided to implement it in their local context. If budgets were used, interventions in areas with high prices would be automatically disadvantaged in any economic comparisons.

The ingredients costing method is generally applicable across all types of educational interventions. However, for SEL interventions the critical issue is how to specify incremental costs relative to the business as usual provision. Where an SEL intervention is delivered as an after-school program, then incremental costs are straightforwardly measured as the cost of that program. However, if the within-school curriculum is altered to emphasize SEL, it may be difficult to identify what incremental resources are involved. More importantly in this case, curricular changes that emphasize SE skills ought to enhance SE skills – that is their purpose. Logically when time and resources are reallocated to a new goal or focus such as SEL, other outcomes obtain less time and resources and so are impaired. At issue is what skills are being displaced with any reallocation of resources and effort from one focus to a different one.

Critically, the costs of SEL interventions must be measured with reference to the overall goals of the education system as well as the specific SE skills that are being targeted. Those who argue that there is no cost to adding an SE intervention to the regular teaching schedule, and therefore that there are no additional instructional costs or displacement of productive activities, should provide evidence that this can be done. Similarly, the burden of proof is on those who argue that time spent on SE skills will not compromise achievement. Certainly, there is a limit to the number of goals and activities that can be addressed within a given time frame. It might be possible to meet this burden of proof: there may be slack in teacher schedules so that they can do more; or it may be that enhanced SE skills complement cognitive learning such that achievement scores and SE skills move mutually upward; or it may be that SEL can indeed be seamlessly integrated into the existing curriculum. However, these possibilities should be substantiated with evidence. As a general rule, instruction in one area should be assumed to

---

10 These prices are taken from appropriate datasets of statistical agencies (e.g. the Bureau of Labor Statistics or the National Center for Educational Statistics); an extensive catalog of prices of education inputs is available from cbce.org.
displace instruction in another area: the working assumption is that there is no ‘free lunch’ in terms of resource use to boost SE skills (or achievement).

3.3 Shadow Price Methods for Benefits

The benefits of interventions are derived from shadow pricing – placing monetary values on – the interventions’ impacts. A shadow price is based upon the societal willingness to pay for a specific impact (Levin and McEwan, 2001: Chapter 4, p.60-61). Conventionally, market prices indicate willingness to pay. But where market prices are missing or are distorted, then shadow prices for willingness to pay must be derived independently. For example, it does not make sense to conceptualize student aggression as a ‘market’ (with buyers and sellers of aggression) and so there is no easy way to observe a ‘price’ for student aggression. Instead, we have to calculate how much we think society is willing to pay to reduce student aggression by a certain amount. That amount will depend on how important society feels about combating aggression; as a lower bound, the value can be represented by how much is currently spent to combat aggression.

In theory, shadow prices can be calculated for most behaviors that are the target of SEL interventions. The fundamental approach of shadow pricing is to measure these behaviors on a scale and then calculate shadow prices for each notch up or down the scale. There are several different techniques for deriving shadow prices (e.g. the burden method, the hedonic method) and the validity of each technique depends on the particular impact being priced out (Boardman et al., 2011).

In practice, deriving shadow prices for impacts from SEL interventions is complex. For some impacts, approximate shadow prices already exist. Published studies have estimated the shadow prices of an array of social ‘ills’, such as smoking, teenage pregnancy, child abuse, asthma, and being a career criminal (Maynard and Hoffman, 2008; Cohen and Piquero; 2009, Table 12; and Cohen et al., 2010, Table 8); for shadow prices of ADHD during the intervention years, see Jones and Foster (2009). However, as we describe below, these shadow prices have to be adapted to specific population groups or recalibrated to specific impacts before they can be applied to SEL interventions.

For other impacts of SEL, new shadow prices must be derived. Given the existing evidence base and methodological research, it is not as yet possible to estimate shadow prices for all possible impacts. Durlak et al. (2011 Table 2) catalog six categories of impacts arising from SEL interventions: (1) academic performance; (2) SE skills; (3) positive social behavior; (4) conduct problems; (5) emotional distress; and (6) attitudes. For academic performance, it is possible to derive or adapt existing shadow prices (see above). For SE skills, it is possible to calculate or approximate shadow prices in terms of increased earnings (see above). For behaviors and conduct, it is possible to derive shadow prices where these actions have clear economic consequences (e.g. a specific juvenile crime). For emotional distress, it is also possible to derive shadow prices, but typically these impacts are less likely to be expressed in ways that do have obvious economic consequences. For the last category – attitudes – it is very difficult to derive shadow prices (Heckman and Kautz, 2012; Borghans et al., 2008). Fundamentally, shadow pricing can only be applied to behaviors (not attitudes) that can be measured with validity.

One short-cut would be to find a ‘unit shadow price’ for specific SE skills, i.e. an overall willingness to pay for increments in these skills (or a measure that encompasses most of this willingness to pay). Two approaches – the mediation of general SE skills through earnings and education – have been discussed above. Two other possible techniques for deriving a unit shadow price are to look at aggression or the economic burden per ‘high-risk’ youth.
Child aggression is typically measured using the aggregated Behavioral Assessment System for Children (BASC). This system has been used extensively in research and is validated (see Reynolds and Kamphaus, 2002, Chapter 3). Helpfully for shadow pricing, BASC T-scores can be scaled to behaviors as: clinically significant (>70), at-risk (60-70), average (40-60) and below average (<40). Also, there is evidence on how aggression persists over time: studies have shown how aggressive and disruptive behavior in primary school predicts aggressive behavior through middle and high school and whether different trajectories of aggressive and disruptive behavior are associated with a range of negative life outcomes in youth and early adulthood. Thus, in theory BASC shadow price notches could be calculated. However, we are not aware of studies that have priced out changes in BASC T-scores either in terms of contemporaneous or lagged benefits. Also, studies typically disaggregate the BASC according to the teacher or child reporter; and it is less clear how the different scores of these reporters independently translate into behaviors and, so, into shadow prices (Reynolds and Kamphaus, 2002). For future research we believe this approach is promising, although it would require a dedicated research endeavor to establish accurate shadow prices for the BASC scale (and so is beyond the scope of our current endeavor).

Another option is to derive a unit value for general youth delinquency that subsequently leads to adult delinquency. For example, Cohen and Piquero (2009, Table 12) have calculated the lifetime social burden per career criminal, one-time offender, high school dropout and drug user over the lifetime. However, there are two issues in applying these values to enhancements in SE skills. One concern is that it may not be valid to translate impacts across any of the six SE categories described above into a general youth delinquency status. This translation requires assumptions about how aggressive behavior relates to criminal activity or externalizing behaviors relate to drug use, for instance. A second issue is that this approach focuses only on the long-term consequences of SE skills and not on the immediate impacts, and the latter may be the most salient (particularly if there is uncertainty about fade-out of SE skills). Overall, this approach requires significant extrapolation and inference to produce shadow prices. However, it can serve as an approximation when no other evidence is available and can be applied in sensitivity analysis.

In summary, the use of a single SEL index – in terms of earnings, or mediated through education, or via the BASC, or a general delinquency index – poses a number of methodological challenges. Therefore, our approach is to derive separate shadow prices for each intervention based directly on the impacts reported for that intervention. This approach, which is more research-intensive, ensures that each intervention is evaluated with respect to its intended objectives. As a sensitivity check, however, we apply these more general approaches to trace their implications for SEL.

Currently, it is not possible to determine a shadow price for every impact from each intervention: there is insufficient data and research evidence to do this. This creates a risk that interventions where data on benefits do exist will appear more beneficial than interventions where the evidence base is more sparse. It also carries the related risk that interventions where more outcomes are evaluated will appear more efficient than those where only a few outcomes are investigated. To illustrate the extent to which impacts are captured by shadow pricing, we create benefit ‘maps’ for each intervention. These maps show the outcomes evaluated, the instruments used to measure these outcomes, and whether

---

11 These outcomes include early sexual activity, teenage pregnancy, school dropout and drug abuse (e.g., Montague et al. (2011) uses BASC scores to link behavior in grades 1-5 to outcomes at age 19/20).
12 A similar approach is undertaken by (Lee et al., 2012).
these outcomes can be expressed as shadow prices. The maps show whether extant interventions can – under current research circumstances – be evaluated appropriately using benefit-cost analysis.

Our benefit-cost analyses of six interventions are reported below. These interventions were selected using several evaluation criteria. Primarily, interventions were selected where there was existing evidence of effectiveness based on a methodologically rigorous evaluation. In addition, interventions were selected with the expectation that costs data could be obtained and that benefits could be assigned shadow prices. Interventions operating across a range of dimensions were also selected, so as to illustrate the variation in efforts to enhance social and emotional learning.
4. SEL INTERVENTIONS

4.1 4Rs

Program

The 4Rs Program (Reading, Writing, Respect, & Resolution) focuses on social and emotional learning and literacy development in grades K-5, with an overall goal of ameliorating aggression and violence at an early age. The 4Rs curriculum, specific to each grade, aims to increase pro-social behavior and help students develop cooperative problem-solving skills. At each grade level, there are seven units, each based on one literary work, that highlight themes such as conflict, diversity, and relationships, and the curriculum reinforces those themes through skills practice.

4Rs has been found to be effective at reducing aggression. In a recent longitudinal evaluation of relatively disadvantaged third-grade children, Jones et al. (2011) employed a school-randomized, experimental design across 1,184 children and 146 teachers in 9 treatment and 9 control group public elementary schools in New York City. The evaluation found improvements of 4Rs compared to ‘business-as-usual’ elementary schooling on: student self-reports on hostile attributional bias, aggressive interpersonal negotiation strategies and depression; teacher reports of attention skills, aggressive and socially competent behavior; and on math/reading achievement. We apply the results from this study to calculate the costs and benefits of 4Rs.

Costs

We estimate the costs of replicating this implementation of the 4Rs program. We consider only the costs of the program above and beyond the resources students already receive as part of their regular instruction in school, i.e., we identify the incremental costs of introducing the programs into existing school activities. All cost estimates exclude any costs associated with the conduct of research activities and reflect only program implementation costs throughout the course of the first two years of the evaluation study. Costs are adjusted for inflation using the CPI-U into 2013 prices and national prices are applied. A discount rate of 3.5% is used.

The cost estimates for 4Rs are adapted from Long et al. (2014, Tables 3 and 4). The analysis by Long et al. (2014) used the ingredients method and high quality data to calculate a whole school costs analysis of 4Rs across grades. Long et al. (2014) calculated costs for all three years during which the longitudinal evaluation of the 4Rs program took place (2004-2007). As the impact evaluation only refers to two years of 4Rs, only two years of costs are reported here and national rather than local prices are applied to provide comparable costs of replication.

Personnel accounted for the most significant portion of program costs. This category includes teachers, center administrators, 4Rs staff developers, and consultants. Incremental school principal time and parental time were not included as these were assumed to be trivial per student. Teacher input was counted for instructional delivery of the curriculum, initial training, ongoing training and workshops. Hours of teacher input were collected from the Center’s weekly accounting logs and

---

13 We appreciate communications from Long et al. (2014) to help us adapt their estimates.
14 Long et al. (2014) compiled the ingredients to implement the 4Rs program through review of documents from the evaluation study (e.g. teacher logs on what they did each day in terms of program implementation etc.), extensive research into the budgets and meetings with the program’s director and accountant. The researchers did not interview teachers or principals directly.
teacher time logs. Overall, the program did not require a particular experience level for these teachers. Teachers devoted time to the delivery of the 4Rs curriculum in each year of program participation. Per year, this totaled to a 7-unit, 21–35 lesson (number depending on grade level), literacy-based curriculum devoted to SEL. Students received one 4Rs lesson per week, one hour per lesson. Before beginning 4Rs, classroom teachers received initial training of approximately 25 hours. Teachers also received on-going training, working with 4Rs staff developers or administrators on-site at their school in individual or small group sessions (approximately 12 contacts each year). Teachers also attended workshops throughout the school year.

Teacher cost was estimated using a national average price for an elementary school teacher in 2013. The same rate was applied for teaching time, initial training, ongoing training and workshops. Prices were adjusted to include fringe benefits valued at 29.5% of salary. Annual salaries were divided by an average academic year of 1,260 hours to convert to hourly wages.

Data on other personnel was also taken from Center’s accounting logs and measured in days. Center administrators included the Executive Director and two administrators (one full-time, one part-time). There were seven 4Rs staff developers working on hourly schedules. (The time of consultants for translation services and curriculum development was not included as this was assumed to be a developmental and not an operating cost). The hourly wage rates (including fringe benefits) for administrators and developers were taken from the database catalog of prices at www.cbcse.org.

Facilities space included workshop space and the Center office space. Two facilities were rented out to host 4Rs workshops. For the workshop space, this was measured as a single unit. The cost per unit was represented as a set fee that the facilities charged for rent. This data was collected using the Center’s accounting records. Additionally, school classroom space for instruction of the 4Rs student cohort was costed out. Space estimates were based on class sizes of 26 students receiving one 4Rs lesson per week of one hour per lesson. Classroom space costs are expressed in national average costs.15

The 4Rs program included kits and basic instructional materials (but no computer materials).16 These were costed according to invoice prices. Travel costs for 4Rs personnel were included based on invoiced amounts. Other subsistence ingredients (such as basic foodstuffs) were also included.

As noted above, a critical feature of costs analysis is that the costs must be measured in relation to outcomes. For 4Rs, this means that we have to decide whether or not to include teaching time as part of the program’s incremental costs. The program was intended to fit into the standard literacy block of the school curriculum in an equivalent classroom space and serve as a partial substitute for regular literacy instruction. The teaching time – and classroom space – is therefore not strictly incremental to the implementation of 4Rs. The only difference was in how these students were taught and so, given the same amount of instructional time, the difference in costs between the methods might therefore be very small. However, the decision as to whether to include instructional time depends on how the impacts of 4Rs are measured. If the impacts that are measured are only those that 4Rs is intended to influence (and not those intended or inherent to the provision received by the control group), then it

---

15 For educational facilities rental rates are not generally available as national prices. We apply school building construction costs (adjusted for cost of land, development, furnishings and equipment) amortized over 30 years. Costs associated with initial training for 4Rs program are amortized over 3 years. We do not amortize ongoing coaching and support to teachers. For classroom space we use new elementary school national median prices amortized over 30 years at a 3% interest rate. Class size estimates are from http://schools.nyc.gov/AboutUs/data/classsize/classsize_2013_11_15.htm. The school year is assumed to be 36 weeks.

16 We amortize initial training and initial materials (4Rs Kits) over 3 years at an interest rate of 3.5%.
## Table 1
Costs for 4Rs Program (Single Cohort for Two Years)

<table>
<thead>
<tr>
<th>Categories/Ingredients</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>630</td>
<td>630</td>
<td>630</td>
</tr>
<tr>
<td><strong>Personnel:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training and support</td>
<td>50%</td>
<td>66%</td>
<td>0%</td>
</tr>
<tr>
<td>Teachers - 4Rs delivery</td>
<td>41%</td>
<td>27%</td>
<td>33%</td>
</tr>
<tr>
<td><strong>Facilities:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rental Morningside office</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>Rental Workshop Space</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Classrooms</td>
<td>7%</td>
<td>5%</td>
<td>6%</td>
</tr>
<tr>
<td><strong>Materials and equipment:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4Rs Kits</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>Supplemental materials</td>
<td>0%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

would not be surprising to see a big impact from the program. Therefore, we report resource costs both with and without instructional time (and the related instructional facilities costs). Including instructional time is conservative, suggesting that 4Rs only influences SE skills and not other skills. Excluding instructional time is possibly too restrictive in that it implies that 4Rs manages to enhance SE skills whilst also maintaining all other skills at the same standard as the business-as-usual approach.

For the two-year program providing 4Rs to a cohort of 630 participants, the present value total cost is $262,300 if instructional costs are excluded or $426,600 if instructional costs are included. Hence, the average cost without instructional time/facilities included is $420; and with instructional time/facilities, the average cost is $680.

### Benefits

There are many impacts from 4Rs that could potentially be translated into money benefits. Based on the available evaluations, we have generated a map of these benefits (see Appendix I, Map 1). There are impacts in: understanding and handling feelings; achievement; academic skills; and other domains. These impacts have specific outcomes that are measurable, but only a subset of these outcomes can be monetized at present. In addition, 4Rs has impacts that are not easily measured (e.g. cooperation or dealing with diversity). Although these impacts are valuable in their own right, if they cannot be measured then it is not possible to assign shadow prices to them.

In this benefit-cost analysis we are able to consider three behaviors. To correspond with the costs analysis, we measure these behaviors at the end of two years of the intervention, i.e. in spring of fourth grade. Our focus is on teacher reports of ADHD symptoms, social competence, and aggressive

---

17 In particular, we are not able to consider the general class of impacts measured using the BASC such as child attitudes and thoughts, even as these may predict subsequent behavior.
behavior. The results based on Jones et al. (2011, Table 3) yield statistically significant effect size impacts after two years for these behaviors.\textsuperscript{18} These effect size gains are: 0.12, for reduced ADHD symptoms; 0.14, for social competence; 0.13, for reduced aggression (average from 0.05-0.21). These effect size gains are equivalent to moving from the median position to the 44th-45th percentile of outcomes.

Shadow prices for ADHD, social competence, and aggression are derived using the cost-of-illness method. That is, the shadow prices are based on what society currently spends on these conditions through the health care system. For the baseline estimates, the ratchet effect is assumed at zero and decay rate is assumed to be infinity. That is, there are no effects in the first year of the intervention, and there are no effects that exist beyond the implementation of the intervention; there are only effects in the second year of the intervention. For ADHD, spending is from Jones et al. (2009). For social competence and aggression, spending on oppositional defiant disorder (ODD) and conduct disorder (CD) respectively are applied; these are from Foster et al. (2005).\textsuperscript{19} These spending figures are for a similarly disadvantaged population to the 4Rs group. All these estimates are conservative in that the ‘cost-of-illness’ method typically excludes some important costs (e.g. family expenditures).

The annual present value benefits of moving from the median burden to the 44th-45th percentile burden are: $2,490 for ADHD; $1,360 for Social Competence; and $4,470 for aggression. Therefore, the present value sum of benefits, i.e. the total immediate benefits of 4Rs expressed as a present value back to the first year of program delivery, is $8,320.

We perform several sensitivity tests based on different assumptions. First, we assume that the ratchet effect is linear. That is the benefits after one year are half the size of the benefits after two years. This assumption is as plausible as assuming there is no effect until after two years of the intervention (but it is less conservative). With a linear ratchet effect the total benefits are $12,630. Second, we apply a decay rate from Washburn et al. (2011). This decay rate is one-third within two years after the cessation of the intervention (and one-third more for the subsequent year). With this decay rate, the total benefits are $16,370. Third, we assume that the decline in ADHD will have an impact on labor market outcomes. Fletcher (2013) estimates ADHD to be associated with a fall in labor market attachment of 5%. Using the CPS data from 2009-2013, we calculate average lifetime earnings of $383,100 in present value terms in 4th grade. With an effect size change of 0.12, the value of the change in labor market attachment is $580-$780. With labor market effects included, the total benefits are $8,930. We consider these to be only a portion of the total benefits, that portion that could be identified and measured, and refer to these as the minimum benefits of the intervention.

\textit{Benefit-Cost Results for 4Rs}

The benefit-cost comparisons for 4Rs are given in Table 2. The estimates of costs per participant are $1,410 and $2,590. But to be conservative, the latter cost estimate is used as the baseline. The minimum commensurate benefits are $8,320. Our main concern is whether the minimum estimated benefits exceed the costs per participant, the gain in net present value when costs are deducted from

\textsuperscript{18} Unadjusted descriptive statistics for these behaviors show no difference between intervention and control groups for ‘Child Social Competence’; adverse effects for ‘Child Aggression’; and positive effects for ‘Child ADHD symptoms’ (Jones et al., 2011, Table S1).

\textsuperscript{19} The annual incremental costs per ADHD=(\$4100-$1800) per year (Jones et al., 2008, Figure 1, estimates for youngest year). For Social Competence: annual incremental costs per ODD=(\$2700-$1490) per year (Foster et al., 2005, Table 1). For Aggression: annual incremental costs per CD=(\$5630-$1490) per year (Foster et al., 2005, Table 1). These figures are in 2000 dollars rounded to nearest ten. Moving down from the median to the 45th percentile is associated with costs that are 3% lower per percentile. The discount rate is 3.5%.
the benefits. The net present value gain at baseline is $5,730 per participant for 4Rs a large gain in benefits relative to costs. We remind the reader that our focus is on whether the minimum estimated benefits exceed the costs, not comparisons with the other interventions, given that the proportion of overall benefits for each intervention captured in the estimates can differ substantially.

The estimates for the net present value for delivering 4Rs to 100 students are given in Table 2. To be conservative, the cost estimate including instructional time is used as the baseline. Therefore, the net present value gain per 100 students is $764,000.

<table>
<thead>
<tr>
<th>Baseline Estimate</th>
<th>Costs</th>
<th>Benefits</th>
<th>Net Present Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(AC2 incl. instruction time)</td>
<td>$68,000</td>
<td>$832,000</td>
<td>$764,000</td>
</tr>
</tbody>
</table>

Sensitivity tests:

- AC2; B + labor market gains | $68,000 | $89,300 | $8,250 |
- AC1 (excl. instruction) | $42,000 | $832,000 | $790,000 |
- AC1; B + labor market gains | $42,000 | $893,000 | $851,000 |
- AC2; ratchet=0.5 | $68,000 | $1,263,000 | $1,195,000 |
- AC1; ratchet=0.5 | $42,000 | $1,263,000 | $1,221,000 |
- AC2; decay rate=0.33 | $68,000 | $1,637,000 | $1,569,000 |
- AC1; decay rate=0.33 | $42,000 | $1,637,000 | $1,595,000 |
- Average across sensitivity tests | $55,000 | $1,203,000 | $1,148,000 |

Sources: Table 1 above. Notes: Present values at third grade (d=3.5%). Baseline model assumes AC2 and baseline benefits. Sensitivity tests sorted by NPV.

To ascertain how robust are the gains in minimum benefits relative to costs, we test economic assumptions that might alter the net returns in Table 2. These alternatives apply the lower cost estimate (excluding instructional time), apply different assumptions about the fade-out and decay rate of effects, and include labor market benefits. Given the conservative assumptions for the baseline estimate, all the sensitivity tests yield higher net present values than the baseline. Across the seven tests, the average net present value is estimated at $1,148,000 per 100 participants. (Moreover, even these results do not include many of the possible impacts from 4Rs itemized in the benefits map). Finally, given the baseline costs of $680, a threshold benefit level can be identified: that is, the economic value of 4Rs is positive even if only the benefits of reduced aggression are accounted for ($4,470); and the value is positive even if only the benefits for ADHD are accounted for ($2,490). Thus, programs such as 4Rs may be economically justified based on very partial measures of benefits in relation to full costs.

### 4.2 Positive Action

Positive Action is a school-based curriculum and supplemental set of school cultural and family activities designed to promote students’ positive thinking, actions, and self-concept. This intervention is
intended to increase intrinsic motivation to learn and reinforce positive behavioral choices. It consists of a series of short lessons at each grade level organized into six units: self-concept; positive actions for your body and mind; managing yourself responsibly; treating others the way you would like to be treated; telling yourself the truth; and improving yourself continually. Supplementary materials can be used by school leaders, teachers, counselors, or families to help students apply the main lessons of the program to target student development in specific areas, such as prevention of bullying or drug abuse. The program is based on the theory that positive actions make us feel good about ourselves, and therefore become self-reinforcing over time.  

Positive Action has been extensively studied, notably in two large-scale, multi-year randomized controlled trials. One study took place in 20 elementary schools in Hawaii and evaluated outcomes across a range of academic, behavioral, school quality, sexual activity, substance abuse, and violence outcomes (Snyder et al., 2010, 2012, 2013; Washburn et al., 2011; and Beets et al., 2009). The other experimental evaluation of Positive Action was conducted in Chicago. With random assignment at the school-level, outcomes from Positive Action were evaluated at seven treatment schools over six years of implementation across grades 3-8. A series of studies examined outcomes such as academic achievement, student behavior, bullying, drug use, violence, and mental health (Lewis et al., 2012, 2013ab; Li et al., 2011; and Bavarian et al., 2013).

We perform our benefit-cost analysis using the outcomes from the trial in Chicago. We chose this site for several reasons. It is the most recent trial of Positive Action. Also, the outcomes measured in that trial most closely align with social and emotional learning outcomes discussed above. Specifically, assignment to Positive Action yielded impacts on: student self-reported normative beliefs about aggression, bullying and other violent or disruptive behaviors; parental reports of bullying or conduct problems; and on disciplinary actions based on administrative data (Lewis et al., 2013). There was a 0.38 effect size reduction in self-reported bullying, as well as there were significant impacts on parent-reported bullying, disruptive behaviors, and disciplinary actions.

**Costs**

We estimated costs using the ingredients method based on interviews from two sites that were part of the Chicago program evaluated by Lewis et al. (2013). The interviews were conducted with a teacher and a counselor who served as site coordinators at two elementary schools. We estimate costs using the ingredients method and calculate only those costs that were incremental to the delivery of ‘business-as-usual’ schooling. All costs are estimated in 2013 dollars and expressed in present values with a discount rate of 3.5%. Although randomization occurred at the school level and all students in schools assigned to treatment were eligible to receive Positive Action services, only the initial third grade cohort was eligible to receive Positive Action services.
followed intensively in terms of measurement and implementation support. Yet, while that cohort was intended to receive Positive Action for six years, student attrition was very high such that average length of treatment was only 3.1 years. Therefore, we estimate the costs of 3 years of treatment, expressed as a present value in 3rd grade.

By far the largest program ingredient is personnel time. This includes the reallocation of classroom teacher time to deliver the program’s approximately 140 15-minute lessons. Classroom teacher time was also incurred for lesson preparation, initial and ongoing training, and for implementation support. Other personnel were also involved in the program. A coordinator at each site, often a teacher or counselor, organized the program and provided delivery support to teachers as well as organized regular parent meetings and school-wide assemblies. School principals spent time supporting the program, and parents contributed time attending meetings and occasionally volunteering in classrooms. Also, at one site a school aide served as an interpreter at parent meetings. Positive Action staff and trainee staff at the University of Illinois-Chicago provided training and implementation support.

Estimates for the amount of time each of these groups contributed to the program were made based on interviews with site coordinators. National average prices for each ingredient in terms of wage time were then obtained from relevant databases. Fringe benefit rates of 29.5% for personnel in K-12 education and 28.5% for personnel in higher education were applied. Annual salaries were divided by an average academic year of 1,260 hours to convert to hourly wages.

Other ingredients include facilities for training, meetings, and most importantly classroom space for program delivery. Facilities costs were based on rental rates for classroom space. Also, program materials were costed, including curriculum kits with workbooks and worksheets for teachers and supplementary kits for counselors, school administrators, and families, and stickers as positive reinforcement for students. Prices for program materials were obtained from the Positive Action catalog and amortized over 5 years, based on average teacher turnover, at a 3.5% interest rate.

Finally, the extent to which teacher time devoted to the program should be considered incremental is unclear. Although there are no additional financial outlays – since teachers would be teaching even in the absence of the program – there is an opportunity cost of lost instruction in what would have happened in the absence of the program. To be conservative, we consider all teacher time to be an incremental cost.

The costs of Positive Action at the two sites are given in Table 3. In total, there were 140 participants in the program over this three year period. The total costs for this cohort was $70,710. The average cost per participant was therefore $510.

We undertake two sensitivity tests on costs. For the baseline estimate all students in a school were assumed to receive Positive Action. All school-wide costs, such as coordinator, principal, and trainer

---

24 These included: the Database of Educational Resource Prices at cbcse.org; and data from the Department of Labor, the Bureau of Labor Statistics, the National Center for Education Statistics, the National Education Association, and the Integrated Postsecondary Education Data System.

25 We assumed an average elementary school classroom size to be 900 square feet, and conservatively assumed that meetings and trainings would take place in a similarly sized regular classroom.

26 In the absence of a significant rental market for educational spaces, we estimated the hourly opportunity cost of using educational facilities by amortizing new construction prices over 30 years at a 3.5% interest rate. Elementary school construction prices were obtained from the 16th Annual School Construction Report at http://webspm.com/~/media/457BAA6BC0164B86A4D4E8E5DD696565.pdf.

27 From the interviews we found notable variation in program implementation between the two sites. Program treatment was notably more intensive in one site (and so variable costs were higher). But the site also served a larger student population (so fixed costs per student were lower).
time as well as school culture materials and activities, are divided among six cohorts. However, coordinators at both sites indicated that treatment intensity in non-evaluated cohorts was lower. Therefore, school-wide costs were instead allocated across students in the evaluated cohorts who received intensive treatment. Under this more stringent assumption, the total program cost was $142,200 or $1,020 per student.

A second sensitivity test was based on length of program receipt. The baseline estimates assume an average of 3 years of program receipt. But students in later years of the study could have benefited from increased program effectiveness due to teacher experience or improved overall school climate (as more of their peers had received the program). As a check, we estimate the program costs spread across six years of program delivery. Under this assumption, program costs using the baseline estimates dividing school-wide costs across 6 cohorts would be $131,800, or $940 per student.

### Benefits

Positive Action is likely to have many impacts on social and emotional outcomes. These include improvements in personal behavior, mental health, achievement and academic behaviors, and school climate (see the Benefits Map in the Appendix I). However, not all these impacts can be measured in such a way as to allow them to be monetized. In addition, some outcomes, such as school quality and academic achievement, overlap with one another, and so to count both of them would be likely to double-count the consequences of Positive Action. Again, we warn against comparing benefit-cost outcomes among interventions when the portions of the benefits that are captured by the evaluations and the ability to convert them to benefits are unknown among the alternatives.

For this benefits analysis, we estimate the shadow price of bullying reduction. Estimates of the shadow price for bullying are only approximate. The economic consequences of bullying include: days of missed school (both for the victim and for the suspended or expelled perpetrator); school personnel time to respond to bullying cases; school practices and training programs to mitigate bullying; parental

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Share of Costs (Site A)</th>
<th>Share of Costs (Site B)</th>
<th>Share of Costs (Sum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants (single cohort for 3 years)</td>
<td>80</td>
<td>60</td>
<td>140</td>
</tr>
<tr>
<td>Personnel:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom teacher</td>
<td>74%</td>
<td>59%</td>
<td>69%</td>
</tr>
<tr>
<td>School staff (coordinator, principal, aide)</td>
<td>11%</td>
<td>8%</td>
<td>10%</td>
</tr>
<tr>
<td>Management (implementation support)</td>
<td>4%</td>
<td>10%</td>
<td>6%</td>
</tr>
<tr>
<td>Parent time</td>
<td>2%</td>
<td>8%</td>
<td>4%</td>
</tr>
<tr>
<td>Training (incl. travel/facilities)</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>Facilities:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom space</td>
<td>6%</td>
<td>10%</td>
<td>7%</td>
</tr>
<tr>
<td>Materials and equipment:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program materials</td>
<td>3%</td>
<td>5%</td>
<td>3%</td>
</tr>
<tr>
<td>Total Resource Cost</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

---

The Economic Value of Social and Emotional Learning – 28 –
time to resolve bullying; resources of social services; resources for alternative placement of perpetrators; and justice system expenditures for cases that involve the legal system.

To our knowledge, no study has estimated the benefits of these reductions using a rigorous shadow pricing technique.\(^{28}\) A very lower bound estimate is to apply the number of school absences that are associated with bullying and price these out using society’s willingness to pay for a day of schooling.\(^{29}\) Under this assumption, an effect size reduction of 0.38 in bullying for three years yields 18 fewer days of bullying. We assume these days build cumulatively over the three year period (equivalent to the ratchet effect for other SEL programs).

**Benefit-Cost Results for Positive Action**

The benefit-cost results for Positive Action are reported in Table 4. Measurable benefits exceed costs by a considerable amount.

<table>
<thead>
<tr>
<th>Sensitivity tests:</th>
<th>Costs</th>
<th>Benefits</th>
<th>Net Present Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>$51,000</td>
<td>$258,000</td>
<td>$207,000</td>
</tr>
<tr>
<td>S1. Zero ratchet effect</td>
<td>$51,000</td>
<td>$810,000</td>
<td>$759,000</td>
</tr>
<tr>
<td>S2. Slower fade-out</td>
<td>$51,000</td>
<td>$392,000</td>
<td>$341,000</td>
</tr>
<tr>
<td>S3. Education benefits</td>
<td>$51,000</td>
<td>$337,000</td>
<td>$286,000</td>
</tr>
<tr>
<td>S4. Delinquent youth benefits</td>
<td>$51,000</td>
<td>$1,007,000</td>
<td>$956,000</td>
</tr>
<tr>
<td>S5. Intensive program</td>
<td>$60,000</td>
<td>$130,000</td>
<td>$70,000</td>
</tr>
<tr>
<td>Average across S1-S5</td>
<td></td>
<td></td>
<td>$443,600</td>
</tr>
</tbody>
</table>

*Sources: Table 3 above. Notes: Present values at third grade (d=3.5%). Sensitivity tests assume costs or benefits from baseline model unless otherwise specified. For delinquent youth benefits, see Appendix Table 1.*

We undertake five sensitivity tests. For the first two tests we vary the ratchet and fade-out effects respectively. For our baseline estimates, we assume that the ratchet effect is cumulative: that is, the benefits accumulate as the intervention is delivered. This assumption is as plausible as linearity (but it is less conservative). Also, for the baseline estimates the fade-out rate is immediate after the final year of evaluation (that is, there are no effects on bullying beyond the length of the program). For our first sensitivity check, we assume the ratchet effect is immediate: that is, the effects of the intervention are immediately equal to the final year impact. Obviously this increases the benefit-cost ratio. As a third

\(^{28}\) For discussions that at least attempts to estimate this issue, see [www.principals.org/Content.aspx?topic=The_Financial_Costs_of_Bullying_Violence_and_Vandalism; www.casafeschools.org/FactSheet5rev2.pdf; and www.highmarkfoundation.org/pdf/publications/HMK_Bullying%20Report_final.pdf]

\(^{29}\) Societal willingness to pay for a school day is estimated using average daily expenditures for a day of public school nationally in 2014 prices. This amount is $70 per day ($13,400/180, see [nces.ed.gov/programs/digest/d12/tables/dt12_215.asp]). There is also a value of increased achievement with fewer days of absenteeism (Gottfried, 2010, 2011ab).
test, we assume that the impacts decline linearly to zero six years after the initiation of the intervention. This too increases the benefit-cost ratio.

The third sensitivity test draws on recent evidence on the long-term effects of bullying reported in Wolke et al. (2013). Although this study is not representative of the U.S. population, it tracks longitudinally students who are bullies, victims, both, and neither. With long-term follow-up of the students, Wolke et al. (2013) identify strong deleterious effects of bullying on all those involved across a range of domains. Specifically, compared to those who have no experience of bullying, victims are 1.1 times as likely to have dropped out of high school, bullies are 1.6 times as likely, and those who are victim/bully are 3.1 times as likely, to drop out of high school. As an approximation, we can use these proportions, along with estimates of the economic burden of high school failure, to derive the shadow price of bullying. Accounting for the impact of Positive Action on bullying, we calculate that the present value benefit per participant from lower rates of bullying is (conservatively) $3,370. These benefits are easily greater than the costs.30

As a fourth sensitivity check we calculate benefits based on impacts reported in Beets et al. (2009) for Positive Action implemented in school in Hawai`i. At this site, the main statistically significant impacts were on substance abuse and violent behavior. For substance abuse, the treatment group rate was 4.0% compared to 7.6% in the control group; for violent behavior the treatment group rate was 6.1% compared to 2.2% (Beets et al., 2009). We use the shadow price of a substance abuser and career criminal from Cohen and Piquero (2009), adjusted to correspond to other shadow prices applied here. This shadow price is then multiplied by the number of reductions in substance abusers or career criminals yielded from a program serving 140 participants. The advantage of this method is that it avoids direct modeling of the lifetime impacts and the fade out of Positive Action. The disadvantage is that it requires extrapolation from behaviors in 5th grade to youth delinquency. Therefore, to be conservative, we assume 20% fade-out each year.31 As well, these shadow prices do not count the burden of delinquency before youth. To avoid double-counting benefits we assume only one of these scenarios applies, i.e. only the substance abuse rate is reduced or the career criminal rate but not both. The calculations of benefits are given in Table 1 of Appendix II. The minimal estimates of benefits exceed costs by considerable amounts. Finally, we assume that the program is delivered as intensively in the second year as it is in the first year. Under this assumption, the average cost of the program rises to $600. Again, the benefits easily exceed these costs.

Overall, the sensitivity analyses indicate that even a partial estimate of benefits for Positive Action exceeds costs across all conditions. Although this analysis yields results that are favorable to Positive Action, the sensitivity tests in Table 4 make clear that the benefit-cost ratio varies widely depending on which model assumptions are made.

### 4.3 Life Skills Training

**Program**

Life Skills Training (LST) is a school-based classroom intervention to reduce substance abuse and violence (www.lifeskillstraining.com). The intervention is generally delivered to at-risk students in middle and/or high school. Life Skills Training teaches social and emotional skills to build confidence and

30 Details of this calculation are available from the authors.
31 This fade-out function approximates to the correlation between 8th grade and adult behaviors of 0.38 reported by Hawkins et al. (1998).
self-esteem, equip youth with the skills needed to resist peer pressure, and generally improve social and emotional competence to reduce anxiety and improve a range of health outcomes. The curriculum is divided into three key components: knowledge and skills needed to resist use of alcohol, tobacco, and other drugs; personal management skills; and general social skills to build assertiveness. The program duration involves sessions over three years, with most of these in the first year of the middle school program followed by booster sessions.

A series of impact evaluations of LST have been conducted (see Botvin et al., 2006; Spoth et al., 2008, Griffin et al., 2011). These studies have examined a range of effects, including reductions in crime, smoking initiation, substance abuse, and risky behaviors (Griffin et al., 2006; Spoth et al., 2008). For each impact, LST is statistically significantly associated with lower delinquency.

Based on these impacts, two benefit-cost analyses of LST have already been conducted (Lee et al., 2012; Miller and Hendrie, 2008). Adapted to be consistent with our other analyses, the results for these two studies are given in Table 2 of Appendix II. The studies differ with respect to estimated costs and how benefits are calculated. In Lee et al. (2012), the benefits are mediated through reductions in crime. In Miller and Hendrie (2008), the benefits are mediated primarily through changes in ‘quality of life’. However, both analyses estimate that the costs of Life Skills Training are very low ($40 and $290 respectively) and that the benefit-cost ratios are extremely large although significantly different (at 46:1 and 21:1 respectively).

Our analysis follows these studies. We describe costs in detail, adjusting for displaced instructional time as appropriate, and itemize the array of possible benefits from LST. We also update the results into 2013 dollars.

Costs

Based on descriptive reports of LST, we calculate the costs by adapting the ingredients identified in the estimates from Miller and Hendrie (2008). Personnel ingredients are based on time spent teaching LST, as well as time on training. Average national teacher salaries are applied, including fringe benefits. Other ingredients include materials and subsistence for teachers. These are priced from several sources. We include costs for personnel time for program delivery.

These costs are given in Table 5. We estimate the cost of three years of LST at $130 per student. These costs are very low because LST is a very light-touch program. However, LST has changed. Earlier evaluations of LST were based on implementation where instruction was within regular class time. In current versions of LST the students are pulled out of class to participate in the program. Thus, we note that these costs may not necessarily represent the same implementation in terms of intensity, training, oversight and support as was used to obtain the published effects used to estimate the benefits.

There has also been a cost-effectiveness analysis of LST compared with other programs that may reduce prescription drug abuse (Crowley et al., 2014).

Ideally, the full ingredients approach should be applied to a version of LST that had been evaluated. However, the identities of recently-evaluated sites were unavailable, and after contacting one of the earlier evaluation sites, we discovered that current program delivery in the school district differs substantially from descriptions of the program as evaluated. Therefore, our cost estimates are based on published descriptions of the program in evaluations. In fact, these cost estimates are roughly in line with those estimated by Miller and Hendrie (2008); they are higher than those of Lee et al. and Aos et al. (2004, 2012) as their estimates of the latter omit personnel costs for program delivery.


These include Western CAPT, Maryland Blueprints, and Aos et al. (2004), as well as the LST website.
Benefits

Our estimate of benefits is derived from impacts reported in Botvin, Griffin, and Nichols (2006). Although LST is primarily a substance abuse prevention program, Botvin, Griffin, and Nichols (2006) estimated the effects of the program on other risky behaviors with the hypothesis that the program could reduce behaviors besides substance abuse through increasing general social and emotional competencies. This study identified a statistically significant reduction in “Delinquency in past year” (odds ratio of 0.684), “Frequent fighting in past year” (odds ratio of 0.742), and “Frequent delinquency in past year” (odds ratio of 0.643) after one year of LST for sixth grade students. Thus, assuming a given population of 100 at-risk students where 50 are delinquent, if those students had received one year of LST, conservatively, only 37 would be delinquent.

Our estimates of benefits follow a similar approach to those for Positive Action. We shadow price the immediate monetary consequences (as the program is only one year, there is no ratchet effect). We then estimate lifetime impacts assuming that these impacts persist. This is likely a lower-bound estimate of the benefits of LST, as delinquency is an ancillary program outcome and because we may not be able to fully capture the economic burden of delinquency.36

At baseline, 53.2% of the combined intervention and control groups displayed any of the above delinquent behaviors, with no statistically significant differences between the groups. Therefore, we can estimate that 1,263 of the 2,374 students in the intervention group had previously exhibited delinquent behaviors, whereas only 937 students in the intervention group reported delinquent behaviors in the follow-up survey three months after the intervention, for a reduction of 326 delinquent students.

As with Positive Action, the immediate economic burden of acts of delinquency can include the value of lost instruction due to absenteeism and truancy, lost instructional time and costs of school staff time devoted to disciplinary actions such as suspensions and expulsions, and the burden experienced

36 As described by Botvin, Griffin, and Nichols (2006), this burden includes “destroying others property, throwing objects at people or cars, shoplifting, stealing from others, taking something from someone by force, or intentionally vandalizing a school or other building.”
by crime victims due to loss or destruction of property. Conservatively, we estimate the value of the reduction in delinquency as a reduction in a single act of the least burdensome delinquent behavior per year for the 326 students whose behavior was positively impacted by the program. This approach is used because the data cannot be disaggregated to show exactly which delinquent acts were avoided and to what extent.

If a large portion of the reduction in delinquent acts is in the most severe acts and for the most serious offenders, the lifetime benefits of delinquency reduction could be significantly larger, ranging up to Cohen and Piquero’s (2009) estimate of $877,030 per career criminal. It is highly unlikely, however, that all of the 53.2% of students identified as exhibiting any delinquent behaviors at baseline could be characterized as “moderate juvenile offenders” or “career criminals,” so we proceed with a more cautious approach of estimating the economic burden of an individual delinquent act. The National Association of Secondary School Principals estimates the loss due to vandalism at $400 per incident and the direct costs of school suspensions at $170 per incident, which translate to $430 and $180 in 2013 dollars, respectively. If 26.8% (1-0.742) of the cases of vandalism, all resulting in suspensions, were averted, LST could save $80 per participant per case per year; if we assume those effects persist without fadeout from the intervention in grade 6 through the end of grade 12, the present value of the benefits are $450 per participant.

**Benefit-Cost Results for LST**

Using our estimates of costs and benefits we can calculate the net benefits for LST, that is the benefits minus costs. This calculation is made within our framework (see above). As noted, the costs of the program are $13,000 for 100 students. With a conservative estimate, the net benefits are $32,000 per 100 participants. Thus, consistent with other studies, we find a strongly positive return to Life Skills Training.

We apply several sensitivity tests to the benefits of Life Skills Training. First, we assume that the only benefit derives from the case of vandalism that is averted per offender: this would reduce the benefits to $80 and yield a negative net present value. For other sensitivity tests, we apply alternative estimates of lifetime costs per at-risk youth: one test applies the benefits per career criminal averted from Cohen and Piquero (2009); another test relies on the association between LST and substance abuse and the benefit of averting such abuse; and a third test applies shadow price estimates from a recent health intervention delivered to 8th grade students to reduce delinquency. Each of these is based upon riskier assumptions for which the evidence on effectiveness of LST is less direct. These sensitivity tests are reported in Table 6.

Overall, Life Skills Training yields a very large present value. Under a range of sensitivity tests, that value may be considerably larger, although we note that these tests are not very precise. Of course, given the very low cost of LST and its targeting of at-risk students, the program does not have to be especially effective in order to generate economic value.

---

38 For the first two tests, we adapt the estimates for career criminals and substance abusers reported in Table 1 of Appendix II. We adjust these amounts for delivery to 10th grade students and for the impacts of LST on being in this status (accounting for fade-out). Conservatively, we report this shadow price assuming the program averts one delinquent per 100 students. Finally, we adapt the shadow prices that Kuklinski et al. (2012) apply in their evaluation of a health intervention to reduce youth deviant behaviors. Adapting their benefits and adjusting them to 2013 prices, the lifetime benefits of $3,100.
4.4 Second Step

Description

Second Step is a classroom-based social skills curriculum for pre-school through the junior year of high school, with a distinct curriculum for each grade. The program builds on cognitive behavioral intervention models and consists of interactive lessons that relate to problem-solving and emotional management. The Second Step program is designed for school-wide implementation and is implemented by classroom teachers. The curriculum is intended to help students develop empathic behaviors and improve their skills in communication, social problem-solving, and critical thinking. The overall goal is for children to identify and understand their emotional state and to manage and communicate these emotions appropriately and so increase social competence and reduce aggressive and delinquent behaviors.

Second Step has been extensively studied. A series of impact evaluations have been conducted in the United States and other countries, and many report positive impacts with students from a variety of age groups, socioeconomic and racial backgrounds, and geographic settings. Between 1989 and 2012 there were 35 studies of Second Step that focused on children with or at-risk of emotional disturbance. However, none of these met What Works Clearinghouse evaluative standards (NCES, 2013). To calculate the costs and benefits of Second Step, we use the impacts from a recent study by Espelage et al. (2013). This study presents first-year results from a 3-year school-randomized controlled trial in 36 Midwestern schools (Chicago/Illinois and Wichita/Kansas). This study utilizes a sample of 6th-grade students who received the 15-lesson sixth grade curriculum. A large proportion of these students were eligible for free or reduced-price lunch (treatment 72.7%, control 75.6%). Thus, the sample is more disadvantaged relative to the national population. Espelage et al. (2013) evaluated the impact of Second Step on reducing youth physical and sexual violence: after one year, the researchers found a substantial reduction in self-reported physical aggression among students.

Table 6

<p>| Benefit-Cost Results for Life Skills Training Per 100 Participants |</p>
<table>
<thead>
<tr>
<th>Costs</th>
<th>Benefits</th>
<th>Net Present Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>$13,000</td>
<td>$45,000</td>
</tr>
<tr>
<td>Sensitivity tests:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1. 1 case vandalism averted</td>
<td>$13,000</td>
<td>$8,000</td>
</tr>
<tr>
<td>S2. Career criminal probability</td>
<td>$13,000</td>
<td>$201,000</td>
</tr>
<tr>
<td>S3. Substance abuse reduction</td>
<td>$13,000</td>
<td>$650,000</td>
</tr>
<tr>
<td>S4. Delinquency reduction</td>
<td>$13,000</td>
<td>$310,000</td>
</tr>
<tr>
<td>Average across S1-S4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sources: Table 1 above. Notes: Present values at third grade (d=3.5%). Baseline model assumes baseline costs and baseline benefits. Sensitivity tests assume costs or benefits from baseline model unless other specified.

Studies have examined a range of effects, including: social competence, school performance, and satisfaction with life (Holsen et al., 2008); positive approach, caring, suppression of aggression, and consideration of others (Cooke et al., 2007); knowledge about empathy, anger management, impulse control, and bully-proofing (Edwards et al., 2005); decline in anxious and depressed behavior (Schick et al., 2003); and aggression and cooperation (Frey et al. 2005).
We estimate the costs of replicating this implementation of the Second Step program based primarily on descriptions of the program.\textsuperscript{40} We consider only the costs of the program above and beyond the resources students already receive as part of their regular instruction in school, i.e., we identify the incremental costs of introducing the programs into existing school activities. All cost estimates exclude any costs associated with the conduct of research activities and reflect only program implementation costs throughout the course of the first year of implementation.

Without access to our usual sources for compiling and clarifying data on ingredients, we had to derive the costs of Second Step using a top-down approach. See Table 7. As described in Espelage et al. (2013), Second Step requires four hours of training, and teachers also completed implementation logs to see what parts of the curriculum had been implemented. Given hourly rates for teacher salaries as well as trainer costs, we estimate the cost of Second Step implementation at $50 per year per student when instructional time is not accounted for.\textsuperscript{41}

In addition, Second Step requires instructional time. Again as described in Espelage et al. (2013), Second Step involves up to 25 instructional hours over the course of an academic year (6th grade). A lower-bound estimate of total instructional hours available per school year is 800, so Second Step absorbs about 3% of total instructional time (or 2% of the entire time in school). Given school expenditures of $12,500 per student, this amounts to $390 for delivery of Second Step on the assumption that it is an addition to the regular school instruction. Overall, therefore, the total resource requirement for Second Step, including instructional costs, is $440. Thus, the critical issue for resource use for Second

\begin{table}
\centering
\caption{Costs for Second Step (1-Year Program)}
\begin{tabular}{|l|c|}
\hline
\textbf{Ingredient} & \textbf{Share of Costs} \\
\hline
Participants & 1,940 \\
Teacher resource for training/logs (including facilities, materials) & 10\% \\
Trainer resource (including facilities, materials) & 1\% \\
Instructional time (incl. classroom space) & 89\% \\
\textbf{Total Cost} (excl. instructional time) & $92,700 \\
\textbf{Total Cost} (incl. instructional time) & $849,300 \\
\textbf{Average Cost per student} (TC1/n) & $50 \\
\textbf{Average Cost per student} (TC2/n) & $440 \\
\hline
\end{tabular}
\end{table}

Notes: Present value ($d=3.5\%$) in 2013 dollars.

\textsuperscript{40}Normally we use informed observers to assist in identifying the required ingredients and their purpose and implementation. Despite repeated attempts within the Chicago Public School system, we were unable to find personnel who had implemented Second Step and had direct information on ingredients usage. A representative of the Office of Social and Emotional Learning at CPS was helpful in confirming some details on the general implementation of the program and how it has evolved over time.\textsuperscript{41} With hourly annual teacher pay at $55,190 (+29.5\% benefits and multiplied by 70\% for overheads, facilities, and management), contact hours are 1,200 per year. Class size is assumed to be 20-24 students. Training is needed annually. Completion of implementation logs requires 2.5 hours annually. Trainer costs are based on salaries of $90,000 (with the same benefits, overheads rates, and class sizes; but contact hours of 600 per year). Materials costs are estimated at $20 per student.

The Economic Value of Social and Emotional Learning
Step is the extent to which it displaces or substitutes for the business-as-usual curriculum. If Second Step displaces current practice it will have a much lower cost than if it is a supplement to the regular curriculum.

**Benefits**

The array of outcomes and benefits of Second Step are shown in the benefits map (Appendix I). These outcomes encompass substance abuse, delinquency, and changes in sexual health status.

Potentially, where these outcomes translate into behaviors, they can also be assessed in monetary values. However, the extent to which these outcomes overlap or confound each other is unknown. As well, it is not possible to tell whether these outcomes are at the expense of achievement gains. Therefore, we apply two approaches to estimating benefits. One is to focus on a single, specific behavior for which plausible shadow prices exist. (This excludes drug knowledge and attitudes but includes sexual risk behaviors which may either induce sexually-transmitted diseases or teenage pregnancy). The other is to use an overall ‘delinquent youth’ shadow price that is intended to capture all the outcomes in aggregate.

The first approach is to calculate a shadow price for aggression. The outcomes of Second Step are from Espelage et al. (2013) who identified a 42% reduction in self-reported physical aggression. Using estimates from Foster et al. (2005) for the medical resources required to address cases of aggression, this outcome yields benefits of \$4,320.42

As an alternative, we apply the shadow price of an at-risk youth calculated by Cohen and Piquero (2009). As reported in Appendix II, this shadow price is (conservatively) \$877,030. From a baseline rate of 5% at-risk youth, Second Step would yield 2.1 percentage points fewer substance abusers to a new rate of 2.9% (corresponding to a 42% reduction in the number of at-risk youth). If the program does not exhibit fade-out, Second Step would yield benefits of \$18,420 per participant. This is a very large benefit compared to the costs. More plausibly, we assume fade-out of 10% per annum until adulthood. Nevertheless, this still yields present value benefits of \$7,550 per Second Step participant.

**Benefit-Cost Results for Second Step**

The costs and minimum net benefits for Second Step are reported in Table 8. Assuming that the costs of instruction are counted, the program costs for one year are \$440. The benefits in terms of reduced aggression are \$4,320. The baseline estimate therefore yields a net benefit of \$388,000 per 100 participants.

The sensitivity tests show that Second Step is likely to yield positive net benefits under a range of scenarios. For example, if we apply the benefits from reducing the number of at-risk youth, then the net present value per 100 students increases from the baseline to \$711,000. If we assume fade-out is not immediate (within one year) but is only complete after three years, then the net present value increases even further to \$795,000. Finally, if we assume that achievement and other school outcomes are unchanged – even as the level of aggression is much lower – then the appropriate cost measure should exclude instructional costs. To the degree that a calmer environment reduces the need for discipline and disruption and makes the school safer, it is even conceivable that academic

42 Calculations assume costs per aggressor of \$1,490 in 2000 dollars with 5% of youth population in this category. Fade-out is immediate after the program is completed.
achievement could increase. In this case, the net resource requirement for Second Step is very small. Yet, it seems unlikely either that it is costless to displace instructional time or that instructional efforts in one direction do not jeopardize efforts in another direction. The overall result suggests a good return on an SEL investment under a variety of assumptions.

### 4.5 Responsive Classroom

**Description**

Responsive Classroom is a pedagogic approach that focuses on how teachers both teach and interact with elementary school students. The approach is designed to provide teachers with strategies, structures, practices, and techniques to improve their self-efficacy, to impact student social and emotional, academic, and other non-academic outcomes, and to build a strong school community.

Through two one week sessions, coaching, and published materials, teachers are trained how to incorporate ten key practices into their teaching philosophy and pedagogy. Teachers use the Responsive Classroom approach to provide emotional support and to proactively manage the class. These practices lead to increased student motivation and engagement, which in turn increases academic skill acquisition.

Responsive Classroom has been identified as an effective program based on CASEL criteria. It is integrated with classroom instruction and provides opportunities for students to practice SEL skills. Also, the program provides tools to assess implementation and evaluate student behavior (CASEL, 2013, Table 3). Early research found that the program improved social skills and reduced problem

---

**Table 8**

<table>
<thead>
<tr>
<th>Benefit-Cost for Second Step Per 100 Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costs</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Baseline</td>
</tr>
<tr>
<td>Sensitivity tests:</td>
</tr>
<tr>
<td>S1. At-risk youth</td>
</tr>
<tr>
<td>S2. Slower fade-out</td>
</tr>
<tr>
<td>S3. Unchanged achievement</td>
</tr>
<tr>
<td>Average across S1-S3</td>
</tr>
</tbody>
</table>

Sources: Table 7 above. Notes: Present values at third grade ($d=3.5\%$). Sensitivity tests assume costs or benefits from baseline model unless other specified. For delinquent youth benefits, see Appendix Table 1.

---

43 Responsive Classroom was developed and is provided by the North East Foundation for Children. The program is composed of ten key practices: morning meeting; rule creation; interactive modeling; positive teacher language; logical consequences; guided discovery; academic choice; classroom organization; working with families; and collaborative problem-solving (see https://www.responsiveclassroom.org/).

44 CASEL (Collaborative for Academic, Social, and Emotional Learning) reviews SEL programs annually to identify effective SEL programs that promote student self-awareness, self-control, social awareness, relationship building, and problem solving. The criteria for the review are that the programs: are well-designed and classroom based; must provide repeated opportunities for students to practice new skills and behaviors; must offer training and implementation support; must be evidence-based (CASEL, 2013).
behaviors (Elliot, 1999; 1995; 1993). More recently, a quasi-experimental evaluation of Responsive Classroom found that the program positively impacted academic performance, teacher self-efficacy, children’s social skills, and children’s perception of school (Rimm-Kaufman et al., 2007; Rimm-Kaufman and Chiu, 2007).

For our benefit-cost evaluation we use evidence from a recent randomized control trial to evaluate the impact of Responsive Classroom on academic achievement in reading and math (Rimm-Kaufman et al., 2014). The sample included 24 schools in a large, diverse mid-Atlantic school district. The treatment group consisted of 13 schools and 1,467 students. Students entering third grade in 2008-2009, fourth grade in 2009-2010, and fifth grade in 2010-2011 were included as participants. Overall, the evaluation was not able to detect a statistically significant impact of the program directly on math or reading achievement at the end of fifth grade (Rimm-Kaufman et al., 2014: Table 4). However, there were statistically significant achievement gains when the program was implemented with fidelity. Structural equation modeling analyses were utilized to estimate the mediated effect of the program through fidelity of implementation (and mediated effect by student’s initial ability). Fidelity was related to effect size gains of 0.26 SD in math and 0.30 SD in reading (Rimm-Kaufman et al., 2014: Table 5). For students with low initial math achievement, implementation of Responsive Classroom in treatment schools was related to an increase of 0.89 SD ($p<.01$) in math and 0.52 SD ($p<.05$) in reading (Rimm-Kaufman et al., 2014: Table 6). For initially high achieving math students, implementing Responsive Classroom with high fidelity in treatment schools was related to an increase of 0.49 SD ($p<.01$) in reading (Rimm-Kaufman et al., 2014: Table 6). Therefore, we perform benefit-cost analysis only for the schools in which the program was faithfully implemented, with the express caution that our findings should not be viewed as representative of typical or average schools.

Costs

No prior information on the costs of Responsive Classroom was available. Our approach was to estimate the ingredients needed to replicate the implementation from the evaluation by Rimm-Kaufman et al. (2014). We were able to obtain ingredients data from the evaluation team based on implementation observation data and other records related to the experiment. In addition, supplementary information was collected from the developers regarding the training provided and how the program has evolved over time.

As per our framework, all cost estimates in our main analyses are incremental and exclude any costs associated with the conduct of research activities and reflect only program implementation costs corresponding to the years of the evaluation study. All costs are estimated in 2013 dollars and expressed in present values using a discount rate of 3.5%. Our data are representative of the whole sample of 13 treatment schools. However, we were unable to interview school-level staff. Because this information was not collected in full accordance with the ingredients method, we perform a series of sensitivity tests on costs. In Rimm-Kaufman et al. (2014), randomization occurred at the school level because the program targets the entire school including principals and teachers. The intent is to create an environment

---

45 We are grateful to Sara Rimm-Kaufman and Julia Thomas at the University of Virginia, and Philip Pohlmeyer and the staff at Responsive Classroom for their assistance.
46 In the ideal case, ingredients information should be collected directly from those personnel who are directly delivering the program or from skilled observers.
where teachers can grow together and to provide the support needed to take risks to advance their skills. The program was rolled-out over three years to longitudinally follow one cohort of students from 3rd through 5th grades.

Personnel ingredients represent the largest input category. Over the course of the three years, two one-week summer training sessions were provided to 3rd-5th grade teachers. These training sessions were supplemented with coaching, online support, and workshops provided by the developer. Principals also participated in the summer training sessions and also received ongoing support and coaching from the program developer. The schools that adopted Responsive Classroom also participated in the annual developer’s conference, which provided additional professional development and networking opportunities.

Personnel costs were estimated using national prices available from the U.S. Department of Labor. To estimate a daily rate, the annual median salary of a principal or elementary teacher was divided by 180 days. The hourly rate assumes a seven hour day for teachers and an eight hour day for principals. If teachers on average spend five years teaching, we assume that the training will last for five years and amortize the cost of the time accordingly using a 3.5% rate. NEFC reported that teachers may remain in the profession longer as a result of participating in the program and working in a school that implements the program well. Thus, we include a sensitivity test that assumes teachers spend 10 years teaching.

Facilities ingredients include the space used for training sessions. All of the training sessions were held during the summer in district space, such as classrooms. We utilize the hourly rate for classroom space for all trainings.47 We also include 9.19% of the average student expenditure in 2013, divided by 180 days, as overhead for the use of facilities to cover additional expenses that would not otherwise occurred.48 The daily overhead rate was amortized for five years using a rate of 3.5%.

Materials included books published by the developer that complement the training provided. Teachers and principals received books and a manual during training. Each school received a set of Responsive Classroom books for their library. The costs of these materials were based on their 2013 sticker prices, with a 20% discount for bulk purchase and amortized with a 3.5% rate assuming the life of each book would be five years.49

We were unable to ascertain the degree of displacement of other activities during the school day due to implementing Responsive Classroom. The program is implemented during the regular school day as a set of embedded practices. The evaluators and NEFC reported that the program provides guidance to teachers to improve classroom organization and schedule management. Through improved transitions and time management, the program does not displace any other class time. Because we were not able to obtain access to interview school-level staff to confirm this, we include a sensitivity analysis that incorporates additional costs for implementation time.

Cost information on Responsive Classroom is given in Table 9. The total cost of the program relative to business-as-usual is $1.32 million over three years. Across the 1,467 participants, this yields an average program cost of $900.

47 Please refer to the Positive Action section above for details.
49 Information from individual communication with NEFC.
We conducted four tests of the sensitivity of the costs per student estimate. First, we included costs for the evaluation personnel who observed the program during the study.\footnote{These costs were not included in our main analysis under the assumption that the observations did not result in any feedback for the teachers regarding their performance. The presence of the evaluators could have changed the instructional environment, perhaps through accountability pressures. However, the findings were not shared with the teachers or principals, so this sensitivity test is conservative.} Two researchers (either study coordinators or graduate students) observed each teacher five times in class for 60 minutes. Including the time for these personnel increases the cost for the program by \$30-\$50 per student.

Second, we modified our assumptions about the duration of the effects of the training. In the baseline, we assumed that training would be needed every five years as teachers exit the program. If Responsive Classroom reduces teacher turnover because of positive effects on teacher satisfaction, training may be needed less frequently. If we assume that the average time a teacher remains in the classroom was increased to 10 years, the cost per student would drop by \$100. Third, we re-estimated the costs of the training using a market price for training instead of the prices paid to the developer. In fact, the costs of training by the developer at the time of the evaluation correspond closely to current market prices for general training. The cost estimate is therefore unchanged.

Finally, we included classroom space and the time that teachers spent providing a morning meeting for students during class time.\footnote{These costs were not included in our main analyses because the teacher’s time and the classroom space used during the class day are included in the base cost of schooling provided to students in both the treatment and control groups.} Although these meetings are only one aspect of the program, their implementation may impact the school day in a costly way. The morning meeting is a 20 minute class meeting designed to build community, teach social skills, and reinforce academic skills and is intended to be held daily. If the cost of teacher time and classroom space are included on the basis of this daily meeting during the school year, the cost per student for Responsive Classroom increases by \$1,060 (to \$1,960).

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Share of Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants (single cohort for 3 years)</td>
<td>1467</td>
</tr>
<tr>
<td><strong>Personnel:</strong></td>
<td></td>
</tr>
<tr>
<td>Principals</td>
<td>13%</td>
</tr>
<tr>
<td>Teachers</td>
<td>70%</td>
</tr>
<tr>
<td>NEFC Training</td>
<td>14%</td>
</tr>
<tr>
<td><strong>Facilities:</strong></td>
<td></td>
</tr>
<tr>
<td>Training space</td>
<td>1%</td>
</tr>
<tr>
<td><strong>Materials and equipment:</strong></td>
<td></td>
</tr>
<tr>
<td>Books/manuals</td>
<td>1%</td>
</tr>
<tr>
<td><strong>Total Resource Cost</strong></td>
<td>$1,101,250</td>
</tr>
<tr>
<td><strong>Average Cost per student</strong> (=$TRC/n)</td>
<td>$900</td>
</tr>
</tbody>
</table>

Notes: Present Values at third grade (\(d=3.5\%\)). National prices. 2013 dollars. Amounts rounded to \$10.
For benefits, we estimate the economic value of achievement gains from Responsive Classroom. Although there was no overall program impact, when results are limited to strong implementation of Responsive Classroom, there was an effect size gain by 5th grade of 0.26 in math and 0.3 in reading (Rimm-Kaufman, 2014, Table 5). These fidelity-determined results are applied here under the justification that comparable effects and costs are being applied. That is, we have estimated costs for a faithfully implemented program that can be combined with the corresponding fidelity-adjusted outcomes. There are limitations with this approach. There could be many reasons why a school that implements Responsive Classroom with high fidelity would have greater gains in student achievement. Therefore, we are unable to know if the results are caused by the program or if they are due to some other difference between the schools. We recommend that these results be interpreted with caution, and they may not be generalizable.

The benefits map (Appendix I) shows that, based on evidence from other studies, there are other important outcomes besides achievement. These include understanding and handling of feelings (measured as social skills, social competence, and school perception) and delinquency, as well as corollary outcomes for teachers (self-efficacy and relationship with the child). However, it is not possible to monetize these impacts either because they are either outside the evaluation by Rimm-Kaufman (2014) or because no shadow prices are currently available.

Based on analysis of the association between achievement and earnings, we calculate the present value of earnings gains from a one standard deviation achievement gain of $34,300 in third grade (see Section 2). Therefore, with an effect size change in overall achievement of 0.26 (the lower bound of gains in math and reading), the value of the change in labor market attachment is $8,920. We note that this value assumes that the gains from Responsive Classroom are permanent (until age 18), even though many interventions experience test score fade-out (e.g., Heckman and Kautz, 2012).

Table 10 shows the benefit-cost results for Responsive Classroom on a per participant basis. With costs of $900 and benefits of $8,920, the net present value per 100 students is $802,000.

<table>
<thead>
<tr>
<th>Benefit-Cost Results for Responsive Classroom Per 100 Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>(high fidelity implementation only)</td>
</tr>
<tr>
<td>Costs</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Baseline</td>
</tr>
<tr>
<td>Sensitivity tests:</td>
</tr>
<tr>
<td>S1. Costs include evaluator observations</td>
</tr>
<tr>
<td>S2. Costs with implementation</td>
</tr>
<tr>
<td>S3. Fade-out of 10% pa</td>
</tr>
<tr>
<td>S4. Fade-out of 25% pa</td>
</tr>
<tr>
<td>Average across S1-S5</td>
</tr>
</tbody>
</table>

Sources: Table 9 above. Notes: Present values at third grade (d=3.5%). Baseline model assumes baseline costs and baseline benefits. Sensitivity tests assume costs or benefits from baseline model unless otherwise specified.
Given the very large baseline benefits, we perform sensitivity tests that show lower bounds to Responsive Classroom. Applying the higher cost estimates described above, the net benefits are reduced. Nevertheless, even including the costs of morning meetings, the benefits exceed the costs by $696,000 per 100 participants. The key sensitivity parameter is the fade-out of benefits from achievement gains. If the achievement gains fade-out by 10% per year until age 18, the benefits are reduced to $3,840. When the fade-out function reaches 25% per year, then the benefits are equal to the costs. Thus, for programs of this scale, the break-even fade-out rate is approximately one-quarter. Looking across the four (conservative) sensitivity tests, the net present value per 100 students is $446,500. However, if we evaluate the program based only on the overall average effects on achievement, the costs would outweigh the benefits. Future studies of the effect of SEL programs on achievement outcomes would benefit from measuring SEL outcomes, such as those in Appendix I, rather than limiting them to standardized test performance.

4.6 Social and Emotional Training

Program

For contrast with the previous programs from the U.S., we decided to apply our methods to an extensive program that was highly documented and evaluated in Sweden. Our final benefit-cost analysis is of Social and Emotional Training (SET) for a program implemented and evaluated for Swedish school children. The SET program is a classroom-based intervention for grades 1-9 designed to support students’ cognitive and social and emotional competencies, learning and development. The program has a similar curriculum to the U.S. Providing Alternative Thinking Strategies (PATHS) program (cite).

The SET program has been evaluated using a longitudinal quasi-experimental design at several different time-points, and the findings show positive effects for the program on a number of social and emotional outcomes. The evaluation addressed 41 classes and 52 teachers at two schools implementing SET and 14 classes at one control school located in the same neighborhood. The impacts of the program are primarily focused on social and emotional competencies linked to mental health issues, although reduced drug use was also considered for students in the older grades. Positive impacts from SET were identified both after two and five years of the program (Kimber, Sandell and Bremberg, 2008; Kimber and Sandell, 2009).

To calculate the costs and benefits of the SET program, the impacts from a five-year follow up (Kimber and Sandell, 2009) are used. In a five-year follow up (Kimber, Sandell and Bremberg, 2008) the results showed positive and significant effects on five of seven variables for the SET-students: internalizing problems (d=.56), externalizing problems (d=.42), mastery as reflecting self-efficacy or hopelessness (d=.36), self-image and self-esteem (d=.54), contentment in school (d=.60) and bullying (d=.35). Effect sizes were medium and no relationship was found between the treatment group and the promotion of social skills (d=.07). The cohort size for the intervention was 1,028 youth, although 837 students participated in the evaluation (663 treatment, 174 control).

---

52 In the two-year follow-up, there were impacts in terms of: in grade 3, psychological well-being (d=.95) and ability (d=.56); and in grades 4-9, body image (d=.48), psychological well-being (d=.33), aggressiveness (d=.32), relation with others (d=.33), attention seeking (d=.32) and bullying (d=.39). In addition, there were significant positive effects for the program on alcohol abuse (d=0.26) and drug use (d=0.23).
Costs

No prior information on the costs of the SET program is available. The costs were estimated by using the ingredients method, and information about the ingredients was collected by interviewing the personnel who implemented the program and from information materials on the program (Kimber, 2001). U.S. prices were used.

This evaluation has estimated the total costs for the entire program over the five years, reported in Table 3 of Appendix II. The cost of the program is calculated for the 1,028 students who participated at the outset of the program. The costs are in 2013 prices and a discount rate of 3.5% is used. Both the operating costs such as teacher salaries and administrative costs and capital costs such as rent costs for classrooms, facilities and materials have been estimated.

Personnel categories include teachers, principals, assistant principals, counselors and program developers. Teacher input was counted for curriculum delivery, initial (two full days) training and ongoing training. Hours of teacher training were collected from the program developer who implemented the program. Teachers in grades 1-5 delivered two 45 minute sessions of SET instruction (the program) every week, each year of the program, a total of 40 sessions per year. In grades 6-9, teachers delivered 45 minutes sessions of SET instruction every week, each year of the program, a total of 40 sessions per year. Teachers received initial and ongoing training for 33.5 hours the first year. In the second year some teachers received additional training and, due to teacher turnover, new teachers received training separately. Training was provided by the program counselor. Data on the national average salaries for teachers in compulsory school, special education teachers, counselors, assistant principals and principals has been taken from Statistics Sweden (Statistics Sweden, 2013).

Facilities for the SET program included school space and training space. The two-day initial training was held in an auditorium at one of the schools. The ongoing training was also held in the facilities at the schools by using classrooms or other smaller meeting-rooms. The costs for facilities are based on the yearly rent in 2013 prices that the two schools pay to the municipality.

The material used in the SET program was a manual used by the teachers which was copied at the schools. Students created their own work-books using paper and other materials provided. The price for the paper was collected from several companies selling paper, and an average was calculated. No travel costs were incurred for training.

In Table 11 the costs are presented. For the five-year program, the present value of total cost is $555,260 and the average cost per student is $540. However, if we exclude the cost for teacher instruction time, the total cost is $143,000 and the average cost falls to $140.

We test this cost estimate for sensitivity to the cohort size. Specifically, as the intervention was delivered to a cohort of only 837 (out of the 1,028 intended participants), then the costs may be apportioned across this number of youth. If the cohort size falls to 837, the average cost rises to either $660 or $170.

Benefits

To calculate the benefits we use the impacts on drug use reported for students in grades 7-9 in Kimber et al. (2009). For drug use, students were assessed at three time points and divided into zero/light

---

53 Turnover affected the number of students who participated in the program for the full five years, although the sample of students in the classes remained stable.
54 The costs for classroom were calculated from the 2013 rent costs to a square meter cost per hour. On average, a classroom is 56 square meters and a class size of 25 students.
users and heavier users. For the zero/light group, students receiving the SET treatment reported a decrease in drug use of 5 percentage points (6% to 1%) from year 1 to 5; zero/light students in the control school reported an increase of 6 percentage points (2% to 8%) from year 1 to 5. This difference equates to an effect size gain of \( d = 0.64 \). For the heavier users, there was an increase for SET treatment group from 1% to 5% and for the control group from 1% to 15% (an effect size, \( d = 0.32 \)).\(^{55}\)

Our benefits map for SET shows many other possible benefits from the program (Appendix I). However, these benefits could not be included in this analysis. Some benefits, such as alcohol, smoking and volatile substances abuse, are likely to be directly confounded with our selected measure of substance abuse. Inclusion of these other benefits would therefore lead to double-counting. Other impacts, such as delinquency and behavior measures of mental health, might also be indirectly confounded with drug use even if these impacts could be accurately shadow priced. Finally, none of the array of social competencies that SET promotes (relations with others, etc.) have shadow prices available and therefore cannot be monetized.

For drug use, we base our shadow prices on what society currently spends on these behaviors through the health care, criminal and judicial systems. For this analysis we use the calculations made

\(^{55}\) Nationally, among males [females] age 15-16, 2% [1%] are drug abusers (illicit drug use during the past 30 days). The proportion raises with age, for 17-18 year-olds the rates are 6% and 2% respectively (The National Board of Health and Welfare, 2013).
by Nilsson and Wadeskog (2013); other spending estimates are given by Nilsson and Wadeskog (2008) and The National Board of Health and Welfare (2013). These estimates are conservative in that they do not include certain costs such as individual losses of income.

Across Sweden, the estimated annual burden of drug use was $3.9 billion ($26 billion kronor) in 2011; this amounts to $450 per capita nationally (SOU, 2011). Of this aggregate amount, 42% was indirect losses of production due to sick leave and premature death, 27% was for spending on the criminal justice system, 26% was for health and social care treatments, and the remaining 5% was for insurance and private health care. Per drug user, the present value of the social burden is estimated at $102,920 in 2013 dollars. Given the respective proportions of youth who are drug users in the intervention versus comparison group, there is a net reduction of 0.0982 drug users. This translates into a benefit of $7,510.

**Benefit-Cost Results for SET**

The net benefits for the SET program are presented in Table 12. The baseline estimates per participant are of costs at $540 and benefits at $7,510, yielding a net present value per 100 participants of $697,000. The SET program easily passes a benefit-cost test: the program is relatively inexpensive per participant; it is highly effective for the population of substance abusers; and the economic burden per substance abuser is very large.

As shown in the bottom panel of Table 12, the SET program is unlikely to have a benefit-cost ratio that is less than one. Under the assumption that the cohort is smaller than expected (837 participants instead of 1,028), and therefore program costs are higher, the net present value per 100 participants is $685,000. Counting only the benefits from heavy users, the benefits are only reduced slightly and the net present value per 100 participants is $322,000. Finally, even if the fade-out rate is 60% within the first year, the net present value per 100 participants is still over $200,000. Moreover, the net present value is likely to be even greater than reported here, given that only the public burden of substance abuse is included (and not the private burden).

<table>
<thead>
<tr>
<th>Table 12</th>
<th>Benefit-Cost Results for SET Per 100 Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>Costs</td>
</tr>
<tr>
<td>Sensitivity tests:</td>
<td></td>
</tr>
<tr>
<td>S1. Smaller cohort of students</td>
<td>Costs</td>
</tr>
<tr>
<td>S2. Heavy users only</td>
<td>Costs</td>
</tr>
<tr>
<td>S3. 60% fade-out in year 1</td>
<td>Costs</td>
</tr>
</tbody>
</table>

Sources: Table 11 above. Notes: Present values (d=3.5%) in 2013 dollars.

---

56 Calculation based on 90% light drug usage (1% treatment, 8% control) and 10% heavy drug usage (5% treatment, 15% control). Of the population, 4% are assumed to be drug users at age 17-18 (The National Board of Health and Welfare, 2013)
5. DEVELOPING METHODS AND EVIDENCE FOR ECONOMIC EVALUATIONS OF SEL

The above analyses indicate that SEL interventions can easily pass a benefit-cost test. In fact, the weighted average benefit-cost ratio across all six interventions with prior evidence of effectiveness indicates that identified benefits outweigh the costs by a factor of 11:1, with an average net present value per 100 participants of $618,380. This finding suggests that had we been able to obtain measures and monetary values of the full range of benefits identified in each of the benefit maps in Appendix I, the benefits would exceed costs by even larger magnitudes. However, a broader conclusion from these analyses is that – if economic evaluations are to have an important influence on the development of SE provision – then significant changes in current evaluation practice are needed (see also Belfield and Levin, 2014). These changes are summarized in Box 5.1.

---

**Box 5.1**
Recommendations for Evaluation Practice

**Costs Analysis**
1. Report all intervention inputs
2. Specify how intervention differs from ‘business-as-usual’

**Benefits Analysis**
1. Draw benefit maps
2. Identify which impacts can and cannot be monetized
3. Identify how impacts overlap

**Methodology**
1. Calculate shadow prices for behaviors
2. Examine resource consequences for schools with students with low SE skills

---

A second concern is with the measurement of impacts and hence benefits. Across SE interventions, outcomes are rarely measured in a consistent way. As our case studies illustrate, evaluations vary in how many domains are measured and what scales are used for measurement. This can lead to benefit-cost analyses where simply the number of benefits varies (regardless of their value). It also means that, instead of determining the appropriate way to measure benefits and to apply shadow pricing techniques, the analyst must build each economic case in idiosyncratic fashion. With more classification and clarification of impacts, many more shadow prices could be estimated and more benefit-cost analysis performed. In addition, it is not clear how persistent the outcomes are. Some programs may affect outcomes in a very ‘fast-acting’ way. Some programs may have effects that endure over a long time period. Both types of programs will be much more beneficial than where effects are delayed or transitory. At present, these ratchet and fade-out attributes cannot be identified from the available evidence and it would therefore be preferable to have outcome data measured over a protracted period of time. All these attributes matter in economic evaluations, and without them it is very difficult to make comparisons across programs.
Applying a direct economic perspective, the main concern is that the methods by which impacts can translate into money benefits is underdeveloped. Even across the six broad categories of outcomes applied in reviews there is lack of clarity as to how outcomes overlap or confound. For example, high achievement may cause high social and emotional skills, e.g. by raising self-respect and a sense of control; the opposite association may hold (as argued by Heckman and Kautz, 2012); or they may be mutually determined. Evaluating an intervention in terms of both gains in achievement and social and emotional skills may therefore be double-counting. Also, it is not clear if SEL outcomes can be expressed monotonically as benefits. That is, when two interventions enhance SEL outcomes by 5% and 10% respectively, it cannot necessarily be assumed that the benefits from the latter are twice those of the former intervention. Similarly, the benefits of SEL may vary across the ability spectrum: improving SEL skills for at-risk students may have very different economic consequences from improving SEL skills across the general population.

There are two clear areas where more shadow pricing would help. First, at this time, there is no mechanism for converting the Teacher Rating Scale (TRS) of the BASC into economic metrics. The BASC-TRS is a widely-used, validated scale for child behavior, and yet its economic implications have not been investigated at all to our knowledge. Second, there is limited evidence on the resource implications within schools from low SE skills (especially negative social behavior or student conduct problems). If SE skills are low and create disruptions and learning problems, both schools and public welfare agencies will require more resources: to hire more counselors or social services personnel; to compensate current personnel who are working with more delinquent students; and to cover other related expenses such as injuries, sick pay, insurance payments, and recruitment costs due to increased quits (Pouliakis and Theodossiou, 2013). Itemized thus, it is seems likely that improving SE skills could yield sizeable cost savings. For example, serious behavior problems can result in expensive special education services and retention in grade or referral to other social agencies. But as yet the dollar amounts of these behaviors are unknown.

In addition, existing shadow prices need to be validated. Even in two areas where some progress has been made – on the association between SE skills and earnings or education – the need for substantial further investigation remains. Alternative scales need to be applied to identify the labor market gains from social and emotional skills more precisely and to model how these gains and skills are mutually determined (Heckman et al., 2006). Also, more emphasis should be put on the impact of SE skills on attainment (such as graduation from high school and college) rather than achievement; the economic value of attainment is much more robust than of achievement, not least because of ‘fade-out’ in cognitive gains.

To address these issues, a range of approaches need to be applied. On the costs side, the ingredients method of calculation is straightforward (although we have noted that it is rarely applied to SE interventions). Moreover, there are two important analytical issues. First, few evaluations give descriptions in comparative terms, i.e. how the intervention differs from business-as-usual. If treatment contrast is not well-documented, it is very difficult to calculate incremental costs. The primary example where this difficulty arises is with in-class treatments and the extent to which these are a substitute for, or addition to, regular classroom instruction. Second, retrospective calculation of cost data is very

---

57 Evidence from Goldhaber et al. (2010) shows that these compensating wage differentials may be difficult to estimate in practice using wage equations. An alternative technique using contingent valuation methods has not yet been applied, nor has analysis of school spending patterns according to student behavior.
difficult. Persons with knowledge of how resources are allocated are often unavailable or are unable to describe program ingredients in detail. During our investigation we struggled to find subject interviewees who could describe the interventions which had been evaluated. This might not matter if SE programs were stable such that current versions of the program resemble past versions. Also, it might not matter if SE programs were always improved with fidelity to program design. Unfortunately, neither condition is likely to hold. This makes cost estimation less precise.

On the benefits side, we recommend the use of benefit maps for each evaluation. These maps show which impacts are measured; whether these impacts are measured consistently; and if there is overlap or confounding across impacts. In turn, we can identify which impacts can be assigned shadow prices, which might be assigned shadow prices, and which cannot. Indeed, the benefit maps we have produced thus far highlight differences across evaluations, and these differences are sufficiently large that we caution, strongly, against any comparison of benefit-cost ratios across the field of social and emotional learning when the ability to capture the benefits among interventions is so variable. Also, we recommend more attention to the wider benefits of SEL. Unavoidably given the current evidence, studies focus on the individual-level benefits of SEL. But there is a strong case that SEL benefits are dispersed through a school or community and so economic value should be measured at that level. A better understanding of the economic value of school climate or classroom behavior would help bring this idea to the forefront.58

Overall, there are many caveats in comparing interventions: student groups differ; and a different set of outcomes are measured using different metrics over different time horizons. However, we emphasize that these caveats are not specific to benefit-cost analysis. They apply to all evaluation research that attempts to compare interventions. In fact, benefit-cost analysis serves to demonstrate the full heterogeneity of the research evidence on social and emotional learning, as well as highlight important gaps in the knowledge base. Equally importantly, this inquiry into the costs and benefits of specific SE interventions suggests that many such interventions might easily pass an economic test of a positive return to investment if sufficient and adequate data were available. Given constraints on education funding, passing this test is an important justification for making further commitments to enhance social and emotional learning.

58 We appreciate this suggestion from Damon Jones.
REFERENCES


THE ECONOMIC VALUE OF SOCIAL AND EMOTIONAL LEARNING

– 55 –
### APPENDIX I: BENEFITS MAPS

**Benefits Map: 4Rs**

<table>
<thead>
<tr>
<th>Outcome Categories</th>
<th>Specific Outcomes</th>
<th>Measures</th>
<th>Monetizable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding and handling feelings</td>
<td>Child aggression</td>
<td>SR - teacher</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Child social competence</td>
<td>SR - teacher</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depressive symptoms</td>
<td>SR - student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hostile attributional bias</td>
<td>SR - student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Normative beliefs about aggression</td>
<td>SR - student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggressive and prosocial fantasies</td>
<td>SR - student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggressive interpersonal negotiation</td>
<td>SR - student</td>
<td></td>
</tr>
<tr>
<td>Health-related</td>
<td>Attention skills - Child ADHD symptoms</td>
<td>SR – teacher</td>
<td>Y</td>
</tr>
<tr>
<td>Achievement</td>
<td>Reading</td>
<td>STest</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Math</td>
<td>STest</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Writing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic skills</td>
<td>Academic skills</td>
<td>SR – teacher</td>
<td>P</td>
</tr>
<tr>
<td>Other</td>
<td>Attendance</td>
<td>Records</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Quality of classroom processes</td>
<td>Observations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social– emotional functioning by teacher</td>
<td>SR – teacher</td>
<td></td>
</tr>
<tr>
<td>Cooperation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Building community</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assertiveness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discussion skills</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem-solving</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dealing with diversity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** SR: Self-report; STest: Standardized test; Records: administrative/school records. P: potentially monetizable.
### Benefits Map: Positive Action

<table>
<thead>
<tr>
<th>Outcome Categories</th>
<th>Specific Outcomes</th>
<th>Measures</th>
<th>Monetizable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal Behaviors</td>
<td>Substance abuse</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SR – teacher</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SR – parent</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Violence</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SR – teacher</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SR – parent</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Records</td>
<td></td>
</tr>
<tr>
<td>Beliefs about aggression and violence</td>
<td>SR – student</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Mental Health</td>
<td>Depression</td>
<td>BASC</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Anxiety</td>
<td>BASC</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Sexual activity</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Affect</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Life satisfaction</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td>Academic</td>
<td>Achievement</td>
<td>STest</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Grades</td>
<td>Records</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Retention</td>
<td>Records</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Academic ability/potential</td>
<td>SR – teacher</td>
<td>P</td>
</tr>
<tr>
<td>Academic Behaviors</td>
<td>Student motivation</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SR – teacher</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Attendance</td>
<td>Records</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Work habits and organizational skills</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SR – teacher</td>
<td></td>
</tr>
<tr>
<td>School Climate</td>
<td>School Quality</td>
<td>Surveys</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** SR: Self-report; STest: Standardized test; Records: administrative/school records; Surveys: devised survey instruments; BASC: Behavioral Assessment System for Children. P: potentially monetizable.
## Benefits Map: Second Step

<table>
<thead>
<tr>
<th>Outcome Categories</th>
<th>Specific Outcomes</th>
<th>Measures</th>
<th>Monetizable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substance abuse</td>
<td>Drug use</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Prescription drug abuse</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Alcohol use (as minor)</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Drug knowledge and attitudes</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td>Delinquency</td>
<td>Aggression</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Delinquency</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Risky driving</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td>Health</td>
<td>Sexual risk behaviors</td>
<td>SR – student</td>
<td>Y</td>
</tr>
</tbody>
</table>

## Benefits Map: LST

<table>
<thead>
<tr>
<th>Outcome Categories</th>
<th>Specific Outcomes</th>
<th>Measures</th>
<th>Monetizable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substance abuse</td>
<td>Drug use</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Prescription drug abuse</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Initiation of other illegal drug use</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Frequency of other illegal drug use</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Alcohol use (as minor)</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Smoking</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Drug knowledge and attitudes</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td>Delinquency</td>
<td>Aggression</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Delinquency</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Risky driving</td>
<td>Records</td>
<td>Y</td>
</tr>
<tr>
<td>Health</td>
<td>Sexual risk behaviors</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td>Social and Emotional Skills</td>
<td>Self-efficacy</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Decision-making skills</td>
<td>SR – student</td>
<td></td>
</tr>
</tbody>
</table>

### Benefits Map: Responsive Classroom

<table>
<thead>
<tr>
<th>Outcome Categories</th>
<th>Specific Outcomes</th>
<th>Measures</th>
<th>Monetizable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievement</td>
<td>Reading</td>
<td>STest</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Math</td>
<td>STest</td>
<td>Y</td>
</tr>
<tr>
<td>Understanding and handling feelings</td>
<td>Social skills</td>
<td>SR – Teacher</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social Competence</td>
<td>SR – Teacher</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Perception of school</td>
<td>SR – Student</td>
<td></td>
</tr>
<tr>
<td>Delinquency</td>
<td>Reduction of Problem behavior</td>
<td>SR – Teacher, Parent, Student</td>
<td>P</td>
</tr>
<tr>
<td>Teacher Outcomes</td>
<td>Teacher Self-Efficacy</td>
<td>SR – Teacher</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher-Child Relationship</td>
<td>SR – Teacher</td>
<td></td>
</tr>
</tbody>
</table>

Notes: SR: Self-report; STest: Standardized test; P: potentially monetizable.
### Benefits Map: SET (Swedish)

<table>
<thead>
<tr>
<th>Outcome Categories</th>
<th>Specific Outcomes</th>
<th>Measures</th>
<th>Monetizable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substance abuse</td>
<td>Drugs</td>
<td>SR – student</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Alcohol</td>
<td>SR – student</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Smoking</td>
<td>SR – student</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Volatile substances</td>
<td>SR – student</td>
<td>P</td>
</tr>
<tr>
<td>Delinquency</td>
<td>Aggressiveness</td>
<td>SR – student</td>
<td>P</td>
</tr>
<tr>
<td>Mental health</td>
<td>Psychological well-being</td>
<td>SR – student</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Body image</td>
<td>SR – student</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Bullying</td>
<td>SR – student</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>Attention seeking</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td>Social competence</td>
<td>Talent/ability</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Relation with others</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Internalizing</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Externalizing</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mastery</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ITIA</td>
<td>SR – student</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social skills</td>
<td>SR – student</td>
<td></td>
</tr>
</tbody>
</table>

Notes: SR: Self-report; P: potentially monetizable.
# APPENDIX II

## Appendix Table 1

Adult Lifetime Benefits from Positive Action

<table>
<thead>
<tr>
<th></th>
<th>Control Group</th>
<th>Treatment Group</th>
<th>Reduction from PA</th>
<th>PV Benefit per avoided lifetime profile¹</th>
<th>Total PV Benefit per Participant</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Substance abuser probability:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimate 1</td>
<td>0.076</td>
<td>0.040</td>
<td>0.036</td>
<td>$877,030</td>
<td>$31,570</td>
</tr>
<tr>
<td>Estimate 2</td>
<td>0.188</td>
<td>0.101</td>
<td>0.087</td>
<td>$877,030</td>
<td>$76,300</td>
</tr>
<tr>
<td>Estimate 3</td>
<td>0.053</td>
<td>0.016</td>
<td>0.037</td>
<td>$877,030</td>
<td>$32,450</td>
</tr>
<tr>
<td>Estimate 4</td>
<td>0.041</td>
<td>0.011</td>
<td>0.030</td>
<td>$877,030</td>
<td>$26,310</td>
</tr>
<tr>
<td>Estimate 5</td>
<td>0.035</td>
<td>0.007</td>
<td>0.028</td>
<td>$877,030</td>
<td>$24,560</td>
</tr>
</tbody>
</table>

| **Career criminal probability:** |               |                 |                   |                                         |                                 |
| Estimate 1           | 0.061         | 0.022           | 0.039             | $2,833,490                             | $110,510                        |
| Estimate 2           | 0.074         | 0.028           | 0.046             | $2,833,490                             | $130,340                        |
| Estimate 3           | 0.038         | 0.011           | 0.027             | $2,833,490                             | $76,500                         |
| Estimate 4           | 0.107         | 0.045           | 0.062             | $2,833,490                             | $175,680                        |
| Estimate 5           | 0.054         | 0.013           | 0.041             | $2,833,490                             | $116,170                        |

### Appendix Table 2
Prior Cost-Benefit Analyses of Life Skills Training

<table>
<thead>
<tr>
<th></th>
<th>Lee et al. (2012)</th>
<th>Miller and Hendrie (2008)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Costs per participant</strong></td>
<td>$40</td>
<td>$290</td>
</tr>
<tr>
<td><strong>Benefits:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical, other resource, work loss, quality of life</td>
<td>$5,960</td>
<td></td>
</tr>
<tr>
<td>Crime</td>
<td>$1,660</td>
<td></td>
</tr>
<tr>
<td>Effects of HS graduation and smoking reduction on earnings and health care spending</td>
<td>$190</td>
<td></td>
</tr>
<tr>
<td><strong>Total Benefits per participant</strong></td>
<td>$1,850</td>
<td>$5,960</td>
</tr>
<tr>
<td><strong>B-C ratio</strong></td>
<td>46.3</td>
<td>20.6</td>
</tr>
<tr>
<td><strong>NPV per participant</strong></td>
<td>$1,810</td>
<td>$5,670</td>
</tr>
</tbody>
</table>

**Notes:** Dollar values in 2013 prices, rounded to $10.
## Appendix Table 3
Costs for SET (5-Year Program)

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>1,028</td>
<td>1,028</td>
<td>1,028</td>
<td>1,028</td>
<td>1,028</td>
<td>1,028</td>
</tr>
<tr>
<td><strong>Personnel:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teachers (Training/ongoing coaching)</td>
<td>$57,000</td>
<td>$17,870</td>
<td>$4,710</td>
<td>$4,710</td>
<td>$4,710</td>
<td>$89,000</td>
</tr>
<tr>
<td>Teachers (SET instruction)</td>
<td>$82,450</td>
<td>$82,450</td>
<td>$82,450</td>
<td>$82,450</td>
<td>$82,450</td>
<td>$412,260</td>
</tr>
<tr>
<td>Administration (Training/support/meetings)</td>
<td>$7,270</td>
<td>$4,480</td>
<td>$1,450</td>
<td>$1,450</td>
<td>$1,450</td>
<td>$16,110</td>
</tr>
<tr>
<td><strong>Facilities:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auditorium</td>
<td>$340</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$0</td>
<td>$340</td>
</tr>
<tr>
<td>Workshop space</td>
<td>$170</td>
<td>$100</td>
<td>$100</td>
<td>$100</td>
<td>$100</td>
<td>$570</td>
</tr>
<tr>
<td>Classrooms</td>
<td>$6,070</td>
<td>$6,070</td>
<td>$6,070</td>
<td>$6,070</td>
<td>$6,070</td>
<td>$30,350</td>
</tr>
<tr>
<td><strong>Materials/equipment:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SET Manual</td>
<td>$60</td>
<td>$60</td>
<td>$60</td>
<td>$60</td>
<td>$60</td>
<td>$310</td>
</tr>
<tr>
<td>Supplemental materials</td>
<td>$1,260</td>
<td>$1,260</td>
<td>$1,260</td>
<td>$1,260</td>
<td>$1,260</td>
<td>$6,320</td>
</tr>
<tr>
<td><strong>Total Resource Cost 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(with teacher SET instruction)</td>
<td>$154,630</td>
<td>$112,290</td>
<td>$96,110</td>
<td>$96,110</td>
<td>$96,110</td>
<td>$555,260</td>
</tr>
<tr>
<td><strong>Total Resource Cost 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(without teacher SET instruction)</td>
<td>$72,180</td>
<td>$29,840</td>
<td>$13,660</td>
<td>$13,660</td>
<td>$13,660</td>
<td>$143,000</td>
</tr>
<tr>
<td><strong>Average Cost 1</strong> (=TRC2/1,028)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$540</td>
</tr>
<tr>
<td><strong>Average Cost 2</strong> (=TRC1/1,028)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$140</td>
</tr>
</tbody>
</table>

Notes: Discounted by 3.5% to year 1. Prices in U.S. dollars (2013). Amounts rounded to $10.